European
University
Institute

Werner Maihofer (Ed.)

# Noi si mura

Selected Working Papers
of the European University
Institute

Badia Fiesolana

W. Maihofer (Ed.), Noi si mura

European University Institute
Institut Universitaire Européen
Europäisches Hochschulinstitut
Istituto Universitario Europeo

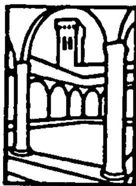
Badia Fiesolana - Firenze

# Noi si mura

Selected Working Papers
of the European University
Institute

Edited by

Werner Maihofer



1986

European University Institute
Badia Fiesolana - Firenze

Office for Official Publications of the European Communities, 1986

© Copyright 1986 by European University Institute, Florence, Italy

*Printed in Belgium*

*à Max Kohnstamm*

# Avant-propos

*Noi si mura,* telle fut la réponse que Cosimo il Vecchio donna à ses amis qui s'enquéraient de savoir pourquoi il se retirait des affaires publiques. Elle exprimait sa conviction ultime que rien ne resterait de lui, que ces bâtiments qu'il construisait alors.

C'est une devise qui aujourd'jui correspond aussi parfaitement à la situation de l'Institut universitaire européen, encore en pleine construction, et installé dans cette merveille de la Renaissance florentine, la Badia Fiesolana, que Cosimo il Vecchio, architecte amateur passionné, bâtit dans les dernières années de sa vie et dont il suivait l'avancement chaque jour sur le chantier.

Nous l'avons donc donnée comme titre à cette sélection de travaux parus ces dernières années, dont le but est d'illustrer comment se bâtit, sans modèle académique préexistant, dans des circonstances parfois difficiles, et non sans périodes d'incertitude, cette communauté scientifique interdisciplinaire et internationale qui a la mission d'initier de jeunes européens à la recherche sur les grands thèmes historiques, juridiques, économiques, sociaux et culturels de l'Europe.

A l'occasion du dixième anniversaire du commencement de cette aventure, nous dédions ce florilège, qui réflète la diversité des conceptions scientifiques et des traditions nationales présentes à l'Institut, à M. Max KOHNSTAMM, qui en fut le premier Président au cours de six années cruciales, et qui a jeté les solides fondations sur lesquelles l'Institut peut s'élever à la hauteur de sa tâche.

*Florence, octobre 1986* Werner Maihofer

## Table of Contents


### I. Political and Social Sciences

JEAN BLONDEL, Essex/Florence
Towards a Comparative Analysis of Cabinet Structures and Decision-Making . . . . . 3

IAN BUDGE, Essex/Florence
Democratic Party Government: Formation and Functioning in 21 Countries . . . . . 17

HANS DAALDER, Leiden/Florence
Parties and Political Mobilization: An Initial Mapping . . . . . 49

BERND MARIN, Vienne/Florence
La recherche en sciences politiques et sociales comme corps de metier: quelques observations comparatives sur le plan international 79

ATHANASIOS MOULAKIS, San Diego/Florenz
Der Souveränitätsbegriff in der Griechischen Verfassung 1975 . 105

BIRGITTA NEDELMANN, Freiburg/Florenz
Geheimnis - Ein interaktionistisches Paradigma . . . . . . . . . 118

JEAN PADIOLEAU, Paris/Florence
Un schéma d'analyse comparative des entreprises de presse . . . 131

PHILIPPE C. SCHMITTER, Stanford/Florence
Democratic Theory and Neo-corporatist Practice . . . . . . . . . 158

### II. Law

MAURO CAPPELLETTI, Florence/Stanford
Repudiating Montesquieu? The Expansion and Legitimacy of "Constitutional Justice" . . . . . 191

TERENCE C. DAINTITH, Dundee/Florence
Legal Analysis of Economic Policy . . . . . . . . . . . . . . . . . 222

BRUNO DE WITTE, Florence
Retour à Costa - La primauté du droit communautaire à la lumière du droit international . . . . . 257

GUNTHER TEUBNER, Bremen/Florenz
Reflexives Recht - Entwicklungsmodelle des Rechts in vergleichender Perspektive . . . . . 290

JOSEPH H. H. WEILER, Ann Arbor/Florence
Supranationalism Revisited - A Retrospective - The European Communities After 30 Years . . . . . 341

## III. Economics

EMIL-MARIA CLAASSEN, Paris/Florence
The Optimum Monetary Constitution: Monetary Integration and Monetary Stability . . . . . . . . . . . . . . . . . . . . . . . . 399

JEAN-PAUL FITOUSSI and DANIEL SZPIRO, Paris/Florence
Adjusting to Competitive Depression: The Case of the Reduction in Working Time . . . . . . . . . . . . . . . . . . . . . . . . 430

WOLFGANG GEBAUER, Frankfurt/Florence
Inflation and Interest: The Fisher Theorem Revisited . . . . . . 446

DOMENICO MARIO NUTI, Birmingham/Florence
Economic Crisis in Eastern Europe: Prospects and Repercussions . 499

JACQUES PELKMANS, Maastricht/Florence
The European Community and the Newly Industrialized Countries 510

MANFRED E. STREIT, Mannheim/Florenz
Heterogene Erwartungen, Preisbildung und Informationseffizienz auf spekulativen Märkten . . . . . . . . . . . . . . . . . . . . . 540

K. VELUPILLAI, Copenhagen/Florence
The Knight And His Utility Function . . . . . . . . . . . . . . 553

## IV. History

GISELA BOCK, Berlin/Florence
Scholars' Wives, Textile Workers and Female Scholars' Work: Historical Perspectives on Working Women's Lives . . . . . . . 569

PETER HERTNER, Darmstadt/Florence
L'industrie électrotechnique allemande entre les deux guerres: à la recherche d'une position internationale perdue . . . . . . . . . . 579

RICHARD T. GRIFFITHS and ALAN MILWARD, Amsterdam & London/Florence
The Beyen Plan and the European Political Community . . . . 596

DANIEL ROCHE, Paris/Florence
Le monde des ombres . . . . . . . . . . . . . . . . . . . . . . . . 623

STUART J. WOOLF, Essex/Florence
Charity and Family Subsistence Florence in the early nineteenth century . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 633

WEINER MAIHOFER, Bielefeld/Florence
Vom republikanischen Revolutionär zum europäischen Patrioten . 648

THE CONTRIBUTORS . . . . . . . . . . . . . . . . . . . . . . . . 663

# I

# Political and Social Sciences

# Towards a Comparative Analysis of Cabinet Structures and Decision-Making

J. BLONDEL
Essex/Florence

In the last decades, serious concern has been expressed, in many Western European countries, about the workings of the cabinet system of government. There have been worries about size, since it is often believed that cabinets of 18, 20, or more, cannot be expected to take their decisions in a truly collective manner; there have been anxieties about the role of prime ministers, whose enhanced status has been felt to contribute to a demotion of other cabinet members; there have been considerable misgivings, in some countries at least, about the rate at which the turnover of ministers, both in general and in a particular job, has been detrimental to their ability to fulfil their role at the head of their departments; and there have been questions relating to the degree of preparedness of ministers to their functions of managers of very large organizations, often dealing with highly technical matters[1]. These matters have been raised to a different extent in various countries, to be sure; indeed the situation is different in each country at least to some degree: but the effectiveness of cabinet government is, to say the least, often being questioned.

Yet this general problem has not so far been tackled on a truly comparative basis. A case is made in some countries, more or less vehemently; documentation and arguments are drawn from national experiences. But it is difficult, if not impossible, to assess to what extent the situation is similar or different from one cabinet to the other; it is not possible to know whether certain practices could be redressed, more or less easily, by borrowing from the practices of other European countries. It is therefore impossible to pinpoint with assurance the precise nature of the difficulties which affect cabinet government.

Meanwhile, we must not forget that the cabinet system has also proved very valuable. It has provided leadership under conditions of emergency, as during World War II in Britain; in normal circumstances, too, it helps to combine the need for guidance by the head of the government with at least a degree of participation by other members of the cabinet. To be sure, this quality may indeed be at the source of the tensions which the

system experiences, as a compromise has to be found between opposite requirements; but this is precisely why the matter needs to be more fully explored, as a balance sheet must be drawn of successes and failures and as the reasons for these successes and failures must be clarified. Such an investigation should be able to show whether a middle course can be steered between possible dangers and what are the conditions under which optimum effectiveness can be attained.

A comparative framework is the natural means by which both the value and the drawbacks of the cabinet system can best be assessed. It is often argued, for instance, at any rate in Britain, that characteristics of modern government in advanced industrial democracies lead inevitably to a move towards 'prime ministerial' government and away from collective decision-making. Manifestly, some trends do exist, such as the 'personalisation' of power resulting from mass media exposure and from the predominant role of party leaders in election campaigns; the technical complexity of departmental decisions, meanwhile, prevents ministers from being meaningfully involved in the affairs of their colleagues. But it is not clear whether these trends exist to the same degree in every European country. Nor is it clear how far other elements, such as the size of the country, or the existence of multi-party coalitions, as well indeed as political traditions in the country do not counterbalance the otherwise general tendency towards 'hierarchical' government.

In order to undertake a comparative analysis, however, a general framework has to be elaborated and a number of reliable indicators have to be found on the basis of which it will be possible to 'locate' individual cabinets in a general space. As a matter of fact, it is probably the difficulty of elaborating satisfactory indicators, rather than the problem of developing a general framework which mostly accounts for the fact that comparative cabinet studies have been slow to progress. But, if progress can be made on both aspects, it will become possible to look for the underlying reasons which account for variations in the character or style of decision-making in parliamentary cabinets. The purpose of this paper is thus not merely to determine what the conceptual framework should be within which to study cabinets, but also to list the variables which are likely to be important and to see to what extent these variables can be operationalized at present.

## I. The General Framework of Analysis

There can be many reasons for analysing cabinets. One aim may be, for instance, to examine the characteristics of national elites and the extent to which they are socially representative of the population or of intermediate

elites. Another may be to examine the extent to which these elites, whether socially representative or not, have views and propose policies which correspond to those of the nation as a whole or of a majority of the nation. Another may be to study coalition formation and maintenance. Yet another may be to study outputs. All these aspects have been, or are studied, indeed occasionally on a comparative basis. Our purpose here is to concentrate more specifically on the cabinet itself: as little is known, as yet, about the ways in which cabinets operate across parliamentary systems, we must map out these characteristics to discover the relationships between the composition of cabinets and the decision-making patterns within them and to assess similarities and differences from one country to another.

In the course of this paper, we shall refer to these cabinets as 'parliamentary cabinets'; as a matter of fact, of course, these cabinets may or may not emanate directly from parliaments: in many cases, civil servants are an important segment of the membership. Moreover, it has been argued occasionally that, with the increase in the power of prime ministers, the expression 'prime ministerial' government is more appropriate; however, given that the extent of 'prime ministerial' dominance is problematic and needs to be ascertained, given also that parliament continues to be the source of political support for Western European cabinets, the expression parliamentary government continues to be justified, whether this support can be expected or not to be given almost automatically as a result of the party configuration in the legislature.

On the other hand, parliamentary *governments* are not necessarily equivalents to parliamentary *cabinets*, as, in many cases, the government is larger than the cabinet. This suggests a hierarchy within the government. There are often, for instance, junior ministers who seem to serve a period of apprenticeship and who do not have a full voice in decision-making. For the present purpose, the investigation is restricted to the cabinet proper, junior ministerial posts being viewed as one of the routes by which politicians can come to the cabinet, a point which may be important in order to determine the extent of specialization of the government.

The main characteristic of parliamentary cabinets is their flexibility. While presidential government is based on the separate existence of an executive and a legislature and attempts to implement, formally, at least, clear-cut distinctions, in parliamentary government, on the contrary, the organising principle is one of conciliation and of linkage. This means, however, that a trade-off has to take place in practice between opposite requirements.

Difficulties arise particularly with respect to two of these requirements. First, there is manifest opposition between *strong leadership and collective decision-making*. Parliamentary government attempts to ensure a certain degree of 'joint' or 'participatory' action at the top; a committee,

rather than a single individual, has the ultimate responsibility; simultaneously, however, parliamentary government is based on the view that it is unrealistic to go too far in the direction of participation. To begin with, the fact that the government, rather than parliament, is in charge of the executive constitutes a concession to the principle of 'leadership', as it implies recognition that only a small number of individuals can be involved in policy implementation and even in policy decision. But parliamentary government goes further: it recognizes that the whole government is unlikely to act effectively if it is truly collective; there has to be 'leadership'. Thus the position of the prime minister is extolled everywhere. The very ambiguity of the old expression *primus inter pares* indicates that parliamentary government is far from rejecting the idea of leadership. The prime minister *is* 'primus'; the chairmanship is not expected to rotate; it is not viewed as insignificant. Thus parliamentary cabinets can be said to be built on an inherent contradiction, which necessarily leads to tension, between the opposite principles of 'collective' action and of 'leadership'. In practice, of course, each parliamentary government will provide a different answer to the way in which this 'contradiction' is to be solved; solutions will indeed vary both with respect to the formal arrangements and to the behavioural patterns; these solutions will also vary over time.

There is tension along a second dimension, however, that of *representation v. administrative implementation or amateurism v. specialisation*. On the one hand, parliamentary government aims at 'implanting' within the upper echelons of the administration men and women who are 'ordinary' politicians, so to speak: ministers are expected 'normally' to be drawn from among the members of the legislature. On the other hand, ministers must also achieve a degree of competence which will enable them to run the government efficiently. As a result, parliamentary government remains basically ambivalent about the relative merits of the 'amateur' and of the 'specialist'. The tension along this dimension may not have emerged in the past as clearly as with respect to the opposition between collective and hierarchical government; but it is manifest at present, both with respect to the background of the ministers and with respect to their relationships with civil servants.

These two types of tensions are inherent in parliamentary government, which has to surmount them if it is to be effective. Consequently, a comprehensive analysis of the characteristics of parliamentary government must be able to provide the means, first, of locating every parliamentary government with respect to these two dimensions and of accounting, second, for the specific locations. This has so far not been the case, largely because a major hurdle to be overcome is that of finding reliable indicators which will enable us to locate a particular cabinet at some point on the two-dimensional space of leadership *v.* collective decision-making and of amateurism *v.* specialization.

## II. The Description of the Characteristics of Cabinets and the Choice of Indicators

Very little has been done so far to attempt to compare the characteristics of national cabinets. Almost all the efforts have been concentrated on the examination of the duration of ministers in the cabinet or in particular posts[2]. Indicators have to be found which will enable us to draw comparisons systematically across time. By and large, the discovery of such indicators is much simpler with respect to the representation-efficiency dimension than with respect to decision-making, both formal (structure) and informal (behaviour). Let us examine these two dimensions in turn.

*The representation-efficiency dimension.* The main problem posed by this dimension consists in evaluating to what extent ministers who occupy a given post can be deemed to have knowledge, experience, and capabilities which are relevant to that post. The qualities which appear to be required are of two types. First, ministers can be deemed to be administratively competent if they are, in a general manner, good administrators or managers; second, they also need to be technically knowledgeable about the subject-matter of their ministry.

Ideally, the indicators chosen should therefore enable us to test whether a given minister is truly able to undertake the task at hand: in practice, we are unlikely to find out, although interviews might help to fill the gap while providing an impression of what ministers aim at achieving[3]. But we can atelast discover how far ministers have been trained to direct the departments at the head of which they are placed by examining their background and their career. First, we can study their *original occupation.* We know, for instance, whether they were civil servants or not, or whether they had a managerial position in the private sector; we also know what their specialization was, for instance whether they were educationalists, farmers, economists, or judges. On the assumption that someone who had a managerial occupation is better prepared for a ministerial job than one who did not have one, we can discover similarities and differences among countries in terms of managerial skills; on the assumption that ministers who are appointed to a post corresponding to their earlier occupation can be deemed to be specialists, we can also discover to what extent cabinets differ in terms of the emphasis given to specialization. A first impression suggests that differences are large. Western Germany, for instance, gives a high premium to specialization, while in Britain and Italy amateurism is widespread[4].

The second type of indicators relates to the careers followed by ministers *after their original occupation and, in particular, during their period in parliament.* Here, too, assumptions must be made. Members of parliaments specialize in certain fields of government, either by inclination or by necessity (for instance, in some countries, the leader of the party will

suggest to colleagues that they must 'look after' a particular area). A large amount of specialization results from the membership of committees and, even more, from committee chairmanships. It seems reasonable to assume that a minister who has been member of a given committee or has been chairman of that committee is likely to have acquired a considerable knowledge of the field and will be more competent to handle the problems of the relevant department. There are of course degrees in the extent of specialization but, by and large, especially if the minister has been member of the committee for a number of years, the assumption about expertise acquired is almost certainly justified up to a point.

The pre-ministerial career is also relevant as a preparation to management, at least in those countries in which there are junior ministers and in which it is common practice for a cabinet minister to have spent previously some time 'under' a full minister. In such cases, it seems reasonable to assume that there is, indeed, a preparation to ministerial office and that ministers have therefore acquired managerial training.

We have thus discovered two broad types of indicators which are likely to suggest that a minister is more or less prepared for the cabinet job which he or she occupies, both in terms of general managerial preparation and specific departmental knowledge. These two modes of preparation are of course not mutually exclusive, although one can substitute for the other. A minister whose first job was in management and/or who became a junior minister is likely to be well-equipped to be an administrator; similarly, a minister whose original occupation was related to the department of the government which he or she subsequently occupied and/or who has specialized in parliament in the field of that department can be said to be prepared to cope with the problems of his/her ministry. It would of course be interesting to discover which, of the original occupation and of the parliamentary career, is the more important for the subsequent ministerial post: this point also needs to be explored. But it is essential to note that the occupational background is not the *only* basis for the specialization of ministers: as Western European ministers, by and large, tend to be drawn from occupational groups which are less directly related to their subsequent governmental office than the ministers of other countries, there is a strong case for examining the training which may be achieved in parliament and in junior departmental positions[5].

A third type of indicator needs to be taken into account: this is the *turnover of ministers* once they are in the cabinet. If a minister holds successively two or more posts, he or she is unlikely to be a specialist in both or all of these fields. It could perhaps happen that the original job specialization and the parliamentary specialization give that minister expert knowledge in two different fields, but it is certainly highly unlikely that such a specialization will be acquired for more than two posts. It follows, therefore, that, where ministers tend to move more from one post

to another, specialization will be lower. Indeed, it can probably be assumed that, in those countries where there is a high propensity for ministers to take different jobs in succession, a higher premium is placed on 'representativeness' than on administrative efficiency[6].

*The leadership-collective action dimension.* The discovery of precise indicators is more complex with respect to the 'leadership-collective action' dimension; indeed, it does not seem possible, so far at least, to discover precise and readily quantifiable indicators. Consequently we need to use indirect means in order to obtain at least a picture of this dimension, which may be thought to constitute the most important aspect of the analysis of cabinets. Indicators are difficult to discover in part because data are hard to obtain: the secrecy of cabinet decision-making processes is well-known and probably universal. But problems arise also for two other reasons. First, structural and behavioural characteristics are often so closely linked, and behavioural patterns modify often only so slowly, so imperceptibly almost, the formal structure that it is difficult to distinguish temporary developments from the 'normal' mode of decision-making; nor is it easy to distinguish underlying trends from freak variations[7]. It is not easy either to discover to what extent a practice started in one cabinet then becomes a rule automatically adopted by the subsequent cabinet. Specifically we have to find whether methods of decision-taking (by voting or consensually, for instance), methods of preparation of decisions (by committees or by individual ministers) are once-and-for-all affairs corresponding, for instance to particular issues debated at a given moment or, on the contrary, constitute the normal way in which business is conducted.

Second, and in contrast with the indicators relating to specialization, many of the indicators which we can use in the field of decision-making are inherently soft; they may enable us to elaborate rough distinctions, typically dichotomies or trichotomies; they do not lend themselves to continuous rankings ranging from fully hierarchical to fully collective. There is a paradox here: while we can conceptually think of a range of situations between total dominance by the leader and wholly collective action, we do not appear to have instruments enabling us to turn this conceptual set of distinctions into a precise ranking.

The published information on the basis of which we can attempt to elaborate these rankings falls under four types. First, we can usually find out which topics reach the cabinet as well as the extent of discussion relating to these topics: we can then conrats these 'cabinet topics' to those which are not considered at the meeting and are decided by ministers individually or by the relevant minister and the prime minister alone. We must expect, however, that the data will not be fully comprehensive. Second, we may be able to obtain information about the extent to which decisions are taken in the cabinet on a consensual or on a majority basis. Such information is almost certainly available for the very critical deci-

sions for which there has been considerable press coverage; but these decisions may only be a small part of the universe and, of course, only if the sample obtained is adequate will it be possible to use this indicator meaningfully. Third, we can also obtain information about the extent to which a long-term strategy is discussed by the cabinet, for instance during special sessions set aside for this purpose. Fourth, we may also discover whether there are periods during which the cabinet discusses, rather informally, the political problems which it faces and the means which will be used to resolve these problems.

These are the only aspects of cabinet life for which we can expect to be able to obtain readily published information. This can be supplemented, to an extent at least, by means of interviews of ex-cabinet ministers, who can provide a general impression of the characteristics of cabinet life. Taken together, these sources of information can provide a significant basis for the discovery of a number of different decision-making styles, both from one country to another and from one cabinet to the next in the same country.

Moreover, alongside data on decision processes in the successive meetings of the cabinets, structural characteristics can help to give an idea of the likely shape of decision-making in general. One such structural characteristic is constituted by the set of distinctions which exist among the ministers (or among the ministries); another results from the arrangements which organize groups of ministers in a particular manner. There are three important sets of such indicators. First, ministers may be hierarchically organized: there may be superministers; there may be ministers-at-large with an overall responsibility; and there may be ministers who depend directly on the prime minister and thus reinforce his or her role. Second, the size and role of the prime minister's office as well as of the ministerial offices (the ministerial *cabinets*) give an indication of the infrastructure at the disposal of each of the members of the cabinet. Third, decisions of the full cabinet may or may not be prepared by committees. As the study by T.T. Mackie and B.W. Hogwood has shown, cabinet committees may be numerous or not; they vary also markedly in importance[8]. One can thus distinguish between permanent and temporary committees, between committees which meet frequently and those which meet on rare occasions, etc. These differences have an impact on the role of the full cabinet and, therefore, on the influence of individual ministers within the cabinet.

A number of indicators do therefore exist which make it possible to begin to describe both the extent of specialization and the nature of the decision-making processes within the cabinets, although there are serious difficulties with respect to both types of fields of study, and especially with respect to the precise monitoring of decision-making characteristics. These analyses will gradually provide a basis for a better understanding of

cabinet life and make it possible to go beyond the general and rather vague comments which are typically made.

## III. The Factors Accounting for Cabinet Structure and Decision-Making: Some Hypotheses to be Tested

We have so far attempted to determine the ways in which the dependent variable, namely cabinet structure and decision-making processes, can be described by means of the indicators which appear to be at our disposal. We need to consider the main factors which can be deemed to constitute the independent variables and account for the ways in which cabinets are structured and take their decisions. Of course, the factors which we are interested in identifying here are what might be called the *proximate factors*, those which can be expected to have a clear and direct effect on the nature of cabinet arrangements. We have to leave aside general considerations about the possible effects of the broader norms of the political system, and, even more, of the socio-economic structure. Whatever influence these factors may be felt to have on the nature of the executive and its shape, the conclusions which might be drawn would remain so vague that it is highly unlikely, to say the least, that any useful conclusions would eventually be drawn from the analysis, beyond, that is, what is commonly said about the general characteristics of political life in Western European societies.

The factors which we need to take into account belong to three main categories. First, a number of technical aspects relating to the composition of the cabinet are likely to structure decision-making directly or, indirectly, because of their effect on the career of ministers. Second some aspects of the political system and in particular of the party system are likely to have a major impact. Third, attitudes and norms relating to leadership create a climate in the nation and in parliament which is likely to influence the relative position of the prime minister and of the ministers.

*Technical factors.* The way the cabinet is composed is likely to affect the character of cabinet decision-making, though the direction, let alone the extent of this effect remains unclear. In the first place, the size of cabinet almost certainly plays a part: the large the cabinet, the less collective the decision process is likely to be, though this relationship is almost certainly not linear; but very small cabinets, in Luxembourg or Iceland are almost certainly different in decision style from larger cabinets, such as those of Britain or Italy. Variations in size are indeed appreciable among Western European countries; there have also been variations over time. If it is indeed the case that larger cabinets are likely to be less collective than smaller cabinets, it would follow that cabinets in the 1980s

are also less collective than their predecessors in the 1940s[9].

Decision-making processes in the cabinet are also likely to be affected by patterns of ministerial direction. It can be hypothesized, for instance, that the role of transient ministers in the cabinet will be more limited than that of ministers of long-standing. If this view is correct, it would further follow that the influence of short-term ministers is particularly small where the average duration of ministers if long and where there are many ministers with a large experience. An index of 'differential duration', based on the ratio of long-standing to transient ministers can be calculated: over the post-World War II period, Ireland, West Germany, Iceland and Sweden have scored high, while Greece, Portugal, Finland and France have scored very low[10].

Third, as we noted already in the context of ministerial specialization, members of governments who hold successively a number of posts can probably be assumed to have more influence over a wide range of issues than ministers who hold only one post. Consequently, in countries or in governments in which there is a higher propensity for ministers to change positions, the collective character is likely to be more pronounced than in countries where ministers customarily hold only one post in the course of their career. If this is the case, and given that ministerial specialization declines as the proportion of ministers who hold successively more than one post increases, the amateurism-specialization dimension appears to be related to an extent at least to the hierarchy-collective decision-making dimension: a government with a high proportion of specialists is unlikely to be markedly collective, though it may not follow that a government with a high proportion of amateurs is likely to be collective.

*The characteristics of the party system.* The party system is likely to affect cabinet decision-making, as well as levels of specialization, because of its effect on the composition of the government. Two aspects of the relationship seem to be particularly important. First, certain *types of coalitions* appear more likely to foster collective decision-making than others and than all types of single party governments. A government composed of an equal number of representatives of three or four parties is less likely to be strongly led by its prime minister than a government which is composed exclusively or even predominantly of members of one party. Three or four party governments may not be truly collective, admittedly: decisions may be taken primarily by those ministers who are regarded as the leaders of the parties in the government; but at least such governments will tend to be 'oligarchical' rather than 'prime ministerial': they have therefore to be located at some distance from the 'hierarchical' end of the continuum of decision-making. It would be valuable to discover whether finer distinctions can be made, for instance between governments composed of two (or one-and-a-half) parties and governments composed of three parties, and between these and governments composed of four

parties or more. Coalitions are also likely to affect, though probably to a limited extent only, the degree of specialization of the ministers. Some parties display a particular interest in some fields of government, whether because of their electoral clientele (farmers) or because of ideological interests (Christian democratic parties with respect to education, possibly); this is likely to create an almost 'natural' relationship between the party and the specific departmental job which, in turn, will result in some degree of specialization. We do not know as yet how widespread the phenomenon is, but given the fact that there is a relatively high level of specialization in a number of continental countries in which coalitions are also prevalent, the extent of the relationship needs to be investigated.

Second, the *internal structure* of parties also affects cabinet structure and decision-making. Parties which are openly factionalized, as is the Christian Democracy in Italy, are not likely to be able or willing to adhere to a highly hierarchical cabinet decision-making style. Parallels can be found in other countries: many, if not most, large Western European parties have 'wings' which often have to be 'represented' in the cabinet and whose members can be expected to wish to have some say in the decision-making process. It seems permissible to hypothesize that, the more the 'factions' or 'wings' of the party in power are clearly defined, the less, other things being equal, it will be possible for the prime minister to take decisions in a hierarchical manner. Factions and wings also have probably some influence on the degree of specialization of ministers, though the direction which this influence takes is not immediately entirely clear: on the one hand, it may be that each faction or wing wishes to control a particular section or field of the government – and specialization may be increased as a result – ; on the other, the desire of each wing to ensure that decision-making is shared relatively equally among the ministers may have the effect of preventing any sector of the government being controlled by the minister concerned either alone or merely in conjunction with the prime minister.

*Leadership norms and characteristics.* Norms relating to leadership influence decision-making styles. Where an emphasis is placed on strong leadership rather than on committee deliberation, cabinet decision-making is likely to be affected. It is, however, difficult to detect the extent to which these norms prevail, let alone the strength of their effect on behaviour. Provisions of the constitution are not a good indicator, as an emphasis on strong leadership in the basic law may simply correspond to a desire on the part of the framers of the document to counteract a prevailing trend towards weak prime ministers: this was to an extent the case in France in 1946, though the tendency was of course more marked in 1958. Conversely, it may be because traditions of leadership are too strong that the costitution emphasizes collective decision-making. There are, however, other possible indicators of the norms relating to leadership. The

constitutions and practices of the parties, for instance, can give a good impression of the position given to chairmen and secretaries. Another indicator is the duration of party leaders in office, as well as the extent to which these are periodically subjected to re-election and have to defend effectively their position[11]. Another set of indicators may be provided by the extent of discretion given to party leaders to take decisions without having to refer to colleagues or to the rank-and-file.

Perhaps the most important discriminating factor might be felt to be, however, the personal characteristics of the prime minister: textbooks and other works emphasize the part played by different prime ministers in ruling 'their' government. It is also commonly believed that there can be considerable variance, within the same country, depending on the personal characteristics of the prime minister: a number of 'charismatic' prime ministers in the post World War II period, appear to have been able to 'rule' their cabinet in a strong, indeed almost autocratic manner. Yet this factor, too, is difficult to operationalize: personality characteristics or 'traits' are rather obscure; no categorization scheme has so far been satisfactorily devised; nor has the measurement of these characteristics advanced markedly[12]. This is a serious handicap, as the study of decision-making processes within cabinets cannot progress significantly until we become able to circumscribe with some measure of precision the possible effect of the personality of leaders on cabinet structure and style. Thus, although it may be necessary to start in a very rough manner and on the basis of very simple distinctions, it is necessary to begin to undertake this task. Perhaps the best is to concentrate on some leaders only; perhaps it is also possible to obtain assessments from members of elite groups about various prime ministers. Only after these preliminary investigations have been conducted will it become possible to determine to what extent and in which circumstances leadership appears to 'make a difference' to cabinet decision-making.

Although these factors do not constitute the totality of the elements which might be taken into account in attempting to 'explain' differences in cabinet structure and decision-making, an exploration into the part which these play will go a long way towards providing clues as to the different modes of behaviour which prevail in Western European cabinets. Moreover, a more precise assessment of the part played by the variables which have been mentioned here is likely to lead to the discovery of other variables, while also providing a basis for the determination of the interrelationship between the various factors. Even if truly precise conclusions cannot easily or quickly be obtained because many indicators are relatively soft, it will be possible to draw conclusions as to the extent to which leadership, party structure, and government composition appear to affect cabinet structure and decision-making. It will thus be possible to clarify a field in which comments are very often made, but little systematic analysis

is undertaken to support these comments. In this way, we will acquire an understanding of the general conditions under which executives have to operate in the context of parliamentary systems.

## Notes

This is a modified version of a paper presented as a position document for a Workshop on Western European Cabinets which was held at Goteborg, Sweden, in April 1986 in the context of the Joint Sessions of the *European Consortium for Political Research*. This paper was designed to serve as a basis for the launching of a comparative project on cabinets taking place at the European University Institute.

[1] These worries were expressed in particular in Britain since the 1960s. R.H.S. Crossman, in his Preface to a new edition of Bagehot's *English Constitution* (1963), London: Fontana/Collins, was particularly concerned about the development of 'prime ministerial government'; P. Kellner and Lord Crowther-Hunt were generally worried about duration and about problems of management in *The Civil Servants*, (1980) London: MacDonald.

[2] Even that literature is still relatively limited. See J. Blondel, *Government Ministers in the contemporary world*, (1985) London and Beverly Hills: Sage; see also V.Herman and J.E. Alt, eds, *Cabinet Studies*, (1975), London: Macmillan.

[3] This was one of the main aims of B. Heady's *British Cabinet Ministers* (1974), London: Allen and Unwin. A comparative analysis is likely to reveal important differences in the extent to which, for example, ministers view themselves as 'policy initiators' or as 'managers', to take two of Heady's categories. J.D. Aberbach, R.D. Putnam and B.A. Rockman, in *Bureaucrats and Politicians in Western Democracies*, (1981), Cambridge,Mass: Harvard U.P., go some way in this direction, but the main aim of the work is to distinguish between the attitudes of ministers in general and those of civil servants.

[4] Between 1945 and 1981, there were as many specialists as non-specialists among West German ministers, while there were six times more 'amateurs' than specialists in the U.K. See J.Blondel, *op.cit.*, Appendix II, p.277.

[5] Western European ministers are less specialized than ministers in other parts of the world and, in particular, than ministers in non-parliamentary executives. See J.Blondel, *op.cit.*, pp. 195 and following.

[6] Britain is a case in point: the movement of ministers from one post to another is relatively high; specialization is relatively low. See J.Blondel, *op.cit.*, Appendix II and note 3, above.

[7] The question of the dominant role of the prime minister is a case in point. Controversies have arisen in Britain as to whether 'prime ministerial government' is a long-term underlying trend, due to profound changes in the characteristics of politics (Crossman's thesis) or merely the result of the vagaries of the personality of individual prime ministers. See R.H.S. Crossman, *op.cit.*, and J.P.Mackintosh, *The British Cabinet*,(1962), London: Stevens.

[8] T.T. Mackie and B.W. Hogwood, *Unlocking the Cabinet*, (1985), London and Beverly Hills: Sage.

[9] Governments have grown everywhere since World War II; in Western Europe, the growth has been smaller than average, but it is none the less substantial, since, on average, they increased by three posts. See J. Blondel, *The Organization of Governments*, (1982),London and Beverly Hills: Sage, pp.181 and following.

[10] The following are the rates of differential duration in the European countries, calculated on the basis of the ratio of ministers having been in office ten years or more (not necessarily continuously) to ministers having been in office one year or less:

| | | | |
|---|---|---|---|
| Austria | 0.71 | Luxembourg | 0.52 |
| Belgium | 0.36 | Malta | No one year |
| Denmark | 0.23 | Netherlands | 0.31 |
| Ireland | 3.80 | Norway | 0.30 |
| Finland | 0.08 | Portugal | 0.12 |
| France | 0.12 | Spain | 0.43 |
| W. Germany | 1.18 | Sweden | 1.00 |
| Greece | 0.03 | Switzerland | No one year |
| Iceland | 1.11 | U.K. | 0.66 |
| Italy | 0.25 | | |

[11] The case of Japanese prime ministers illustrates this point: their tenure in office is markedly influenced by the fact that they have to be re-elected as leaders of the liberal-democratic party periodically.

[12] See in particular B.M. Bass, *Stogdill's Handbook on Leadership*,(1981), New York: Free Press. See also J. Blondel, *Political Leadership*, forthcoming.

# Democratic Party Government: Formation and Functioning in 21 Countries

IAN BUDGE
Essex/Florence

## Introduction

Curiously, the behaviour of parties in elections has attracted more attention over the past three decades than the behaviour of parties in government – at least on a systematic and comparative basis. Yet there can be no argument over the equal if not greater importance of governmental processes for our understanding of politics. No matter how free, representative and responsible elections may be, any failure to respond at government level renders them totally ineffective.

Parties alone operate in elections as well as in government. They are consequently in a unique position to channel broad popular preferences into government action. Studying the way parties do this is just as essential as the analysis of their electoral strategies for an understanding of democratic processes.

This in turn is vital to the justification of democracy against other types of political system. Without valid knowledge of how governments are formed and run by parties, we cannot argue with any conviction for the general superiority of the democratic system. We are simply basing assertions on ignorance of its central processes.

Nor is it enough to review the workings of party governments in one or two sophisticated Western politics. For this, quite fairly, provokes the criticism that democratic parties may govern efficiently and sensitively under favourable conditions, but where they face the social and political traumas of the rest of the world will prove just as unresponsive and ineffective as any other type of ruler. To demonstrate that democracy does live up to its claims, one must study the behaviour of parties in government for as many democracies as possible, comparatively and systematically.

This book reviews party governments in 21 democracies, in the underdeveloped as well as the developed countries of the world, operating under a variety of cultural and societal conditions. These countries are chosen as

having maintained a democratic system for most of the post-war period. Countries which became independent or democratic later are excluded because their party systems and modes of government have in most cases not yet stabilized, and cannot be used as a basis for broad comparative and temporal generalizations. This still permits the inclusion of India and Sri Lanka in the analysis, together with Japan, Australia, New Zealand and Canada, outside the core democracies of Western Europe. As the object is to study countries with broadly similar government arrangements (although operating under a variety of cultural, social and economic circumstances), the United States and Switzerland are also excluded. The presidential government of the United States, where the electorally successful party takes executive powers unconditionally and with a constitutionally guaranteed term of office, differs sharply from the systems operated in the countries chosen for study, where government's tenure of office depends upon its ability to win votes of confidence in the legislature. Similarly, none of the other countries has experienced the radical devolution of power to the cantons made in Switzerland and the consequent relegation of national government to a caretaker role which facilitates permanent coalitions between the main parties.

The criteria of selection nevertheless allow for inclusion of all the other major democracies of the world and thus for the testing of a comprehensive theory of democratic party government. The concern to develop such a theory stems from three considerations. The first is a desire to round off previous work with Dennis Farlie, on the behaviour of parties in elections and the effects of this on voting and election outcomes (Budge, Crewe and Farlie (eds.), 1976; Budge and Farlie, 1977; Budge and Farlie, 1983). In *Explaining and predicting elections* (1983) a systematic theory of election outcomes was developed and validated for 23 democracies – the ones studies here with the United States and Switzerland. Explaining the behaviour of parties in government is a natural corollary to explaining how they gain the popular support necessary to sustain a governmental role. Together, the theories of elections and governments constitute an overall explanation of the central democratic processes.

Work on the theory of government actually began as part of the project on "The future of party government", directed from 1981 by Rudolph Wildenmann at the European University Institute in Florence. Analyses of the ways party roles in government will develop in the future, and indeed descriptions of the way it functions at present, all require a comprehensive theory of what parties do in government. This provides general reference points for detailed analyses of particular processes. In turn, these detailed analyses contribute to the development of an overall theory. So there is a symbiotic relationship between the broad comparative analysis discussed here, and the country of process-specific analyses produced by the project (series references to books from party project).

Not only did theory and book gain intellectually from the wide-ranging seminars organized under the party government programme, but they also gained necessary material support, in the shape of research assistance, photocopying and travel costs – without which the large amount of information needed to validate the theory would not have been assembled and processed. At an early stage the Nuffield Foundation also contributed to these costs in its usual enlightened and timely fashion (Grant No SOC/181/755). I am grateful to all these friends and supporters whose essential help prevents them from disclaiming all responsibility for what follows! In particular I should like to thank Rudolph Wildenmann, Val Herman, Norman Schofield, Dick Katz. While the book has incorporated many of the criticisms and suggestions resulting from discussions with them, I am solely responsible for errors of execution and presentation.

The third impulse to generate a systematic theory of democratic party government stems from the desire to develop and unify work already done by others. While the comparative analysis of democratic governments has been neglected it has not been totally ignored. In particular, attempts have been made to create a theory of the way parties enter into coalitions, and distribute ministries, and to test this comparatively, within the context of an office-seeking model of party behaviour.

It will be obvious that the theory developed here (as any theory in the field must, in my opinion) starts from an evaluation of the strengths and weaknesses of such office-seeking models. While these vary in detail, their basic assumption is that democratic parties aim above all at acquiring and keeping government office. More detailed implications are then derived from this and used to explain actual party behaviour.

Such models have had two enduring effects on subsequent research. One is to ensure that politicians' strategies and decisions are viewed as broadly rational and hence explicable in fairly simple terms. The theory stated in Chapter 1 stands like the office-seeking explanations within the broad "rational choice" tradition, even if it attributes more substantive preferences to party politicians than the abstract maximization hypotheses derived from economics.

The second influence from the "office-seeking" models is apparent in the a priori, semi-deductive presentation of the book's underlying theory. After initial assumptions have been stated, their implications are drawn out before being checked against actual party behaviour to see if they hold. Such a priori theory is useful because it must fill in all the links in its chain of reasoning, and is thus more specific and detailed about all the assumptions involved than most retrospective interpretations of pre-existing data. A fully articulated statement is not only more satisfactory in itself but provides a better basis for prediction of future behaviour. Although prediction is not the same as explanation, it is both of practical use – particularly in directing attention to what is going to happen in such a vital

institution as government – and important in selecting the best theoretical explanation. The systematic and predictive form of the theory also has the advantage of dovetailing with the predictive theory of election outcomes already developed (Budge and Farlie, 1983).

Not that inductive data-based work has been ignored. This book is from one point of view an attempt to synthesize these empirical findings within the framework of rational choice theory. Overwhelmingly, detailed research has shown that party leaders' policy preferences cannot be ignored in any realistic explanation of government processes. Such preferences not only limit what politicians will do and say to get elected (Robertson, 1976; Budge and Farlie, 1977, Chapter 11). They also make it unsatisfactory for them to govern without giving precedence to ideological predispositions and policy commitments (Budge and Farlie, 1977, pp. 157-162; Rallings, 1984). These markedly affect, for example, different governments' expenditure decisions (Castles (ed.), 1982). Subsequent discussion not only incorporates the research-attested primacy of policy as a basic assumption, but uses some of the previous findings to decide between competing theories (Chapter 4 below).

The order of chapters is determined by the nature of the approach. Chapter 1 discusses previous theory and research and criticisms and developments which lead to the new deductive theory. This is summarized in the verbal propositions of Tables 1.1 to 1.3, with the supporting text. Together these constitute a systematic, comparative, comprehensive theory of government formation; initial distribution and re-allocation of ministries; policy-making and outputs; and decease. Chapters 2 to 5 each start by drawing out implications of the unified theory for their own area, and then report the fit between these implications and the actual record of post-war governments in the 21 countries. Chapter 2 is concerned with government formation; Chapter 3 with the initial distribution of ministries between parties; Chapter 4 reviews policies and expenditures; and Chapter 5 deals with reshuffles (often involving internal party factions) and the collapse of governments. Chapter 6 assesses the bearings of the empirical findings on general theory.

It will be obvious from the outline that the theory here, unlike its predecessors, covers all the major preoccupations of parties in government, treating the actual formation of governments as part of their overall life-process rather than as the major problem requiring explanation.

This seems necessary if the explanation is to provide a context for general research into problems of party government. For historians, and institutional and policy analysts, the actual formation of a government is less significant than what it does during its lifetime. Concentration on the emergence of particular types of government is perhaps natural in theories that have emphasized the predominant concern of parties with gaining and keeping office. For other research it is only a starting point.

Besides explaining the processes internal to government formation, allocation and change of ministries, policy-making and termination, a satisfactory theory should also relate them through a concise set of general assumptions linking what would otherwise be regarded as disparate areas of activity. If the assumptions can be shown to fit actual observations, either directly or through their implications, they can be taken as basic principles underlying the whole operation of democratic party governments. Tables 1.1 to 1.3 of Chapter 1 attempt to present such general assumptions succinctly in the shape of verbal propositions which can be precisely discussed and whose relevance to the various implications and applications checked in later chapters can be made very clear. The presentation and discussion of these assumptions, and of their underlying rationale, form the main concerns of this introductory chapter.

## Parties in government – A new theory

### 1. Previous explanations: The primacy of office and the introduction of policy considerations

Before a new theory can be developed, careful attention must be paid to the old. It is rare in political science to have a body of well-developed earlier theory. So some justification must be given of the need to modify it. Moreover, the new theory, like any other systematic explanation of the behaviour of democratic governments, must define its position in regard to the ideas about "rational choice", borrowed from economics, which have dominated studies of politicians' behaviour over the past 25 years. Because variants of such ideas constitute the only general theories of democratic government at the present time, it is essential to come to terms with them before developing any other approach.

This does not mean that substantive research findings made independently of this framework should be ignored. These in fact constitute one of the main reasons for modifying current models and will be introduced into discussion at the appropriate moment. But discussion must begin at the theoretical level.

Office-seeking assumptions are not the only ones which might be postulated within the framework of "rational choice". Rational choice in a broad sense is simply adoption of the most cost-efficient course of action to achieve desired ends. Rationality relates to the pursuit of given ends and cannot of itself impose restrictions on the type of end endorsed (Budge, Farlie and Laver, 1982).

In economics the further restrictive assumption is imposed (as money is so essential to gaining material ends) that maximizing profits and minimizing costs can be taken as a universal immediate objective for everyone in

the market. By analogy, most extant models of party behaviour in government assume that possession of office is so essential to achieving all other party goals, including "altruistic" or ideologically motivated enactment of their policies, that office becomes their universal immediate objective. Parties are therefore seen as trying to hold as many government ministries as they can for as long a period as possible at least cost in terms of effort expended.

While the essence of money is that, once gained, it can be freely spent, office is usually held only on certain conditions. So the analogy is not perfect. What happens where the conditions of holding office preclude the achievement of preferred policy ends is not usually discussed in office-holding models. One analysis indicates that under such circumstances, negative utilities might emerge rather quickly (Budge and Farlie, 1977, pp. 157-162).

This is simply common sense. It is hard to imagine any politician in the real world agreeing to enact policies he opposes simply to stay in office. There are prudential as well as moral reasons for this. Even in a world of selfish utility-maximizers a reputation for responsibility and reliability is worth votes (Downs, 1957, pp. 103-109). Yet if politicians seek office as their immediate overriding objective, they will adopt just such a course. Office is justified as essential to the attainment of other ends, but rather quickly replaces these within the strict office-seeking models (partly because other ends are vague and unspecified).

The basic assumption behind office-seeking developments of rational choice theory is therefore that parties (or, in some variants, factions within parties) are united groups aiming at a maximal share of the spoils of office. They must therefore form a coalition which commands over 50% of legislative seats (otherwise their equally selfish opponents would combine to vote down the government), but not much over 50%, otherwise they would have to share their spoils more widely than would otherwise be required (Riker, 1962; Gamson, 1961). Although it applies mainly to coalition formation in a multi-party situation, and not to other government processes, this theory is general and comparative. It meshes conveniently with models of voting behaviour where parties are similarly assumed to be selfish office-seekers, adjusting policies to get votes (Downs, 1957; for a review see Budge and Farlie, 1977, Chapters 3-5). Combined, the two office-seeking models form a comprehensive and unified explanation of democratic processes (including government policy-making, which would in this case be directed to aggrandizement of party interests and attracting further votes). If one party has a majority in government the situation from the point of view of office-seekers is even better, as they do not need to share their spoils with anyone outside the party.

Various extensions have been proposed to this basic position. Some of these have been along the lines of refining previous requirements, so that

not only a minimal number of legislators but also of parties share the spoils (for a review of such principles see Taylor, 1972). None of these modifications, however, noticeably improves the fit of the original model to actual patterns of coalition formation in Western Europe (Taylor and Laver, 1973; Herman and Pope, 1973; Laver, 1974).

The most important extension to the original assumptions has been to introduce ideology into the rational choice framework, although more as an externally imposed constraint on the coalitions office-seekers can form, than as a modification of that primary urge. The argument, in other words, is that given the existence of ideology it makes sense even for those not ideologically motivated themselves to combine with the closer parties rather than those further away. The reasons stem from the tensions and disagreements which ideological disparities would introduce. These produce higher costs – time and energy spent in internal government negotiations – which in turn diminish the net profit to be made out of office-holding. Moreover internal disagreements also render the fall of the government more likely.

For both reasons policy disagreements are viewed even on office-seeking assumptions as needing to be minimized either by ensuring that ideologically contiguous parties form the government (a minimal connected winning coalition (Axelrod, 1970] or that overall diversity is reduced, even although this may sometimes involve "jumping" small neighbouring parties (Leiderson, 1966; De Swaan, 1973). Such considerations are then assumed to enter increasingly into politicians' calculations, along with their desire to attain and keep office with the smallest group possible (again, of course, a majority party is ideal from this point of view). (For a review of the ways in which these criteria interrelate see Taylor, 1972.)

Such extensions to the original minimal winning criteria can be criticized. If ideology is so important a constraint, why should it, rather than office-seeking, not dominate the actions of politicians? Attainment of office is often postulated as a primary goal because it is the prerequisite to putting policy into practice. If, on the other hand, the attraction of ideology is great, its binding power in a coalition would surely substitute for common pursuit of gains from office, thus making for less concern with a minimalist criterion, as Browne and Rice (1979) and Grofman (1980) imply. Since ideological sympathies would draw together elements of government and opposition, there would also be more prospects of external support (and internal disintegration) than the pure office-seeking theory allows.

Ideological sympathies and alliances fluctuate, as different aspects of policy assume prominence. However, both the original, pure office-seeking theories, and their ideological modifications, are static. They assume that the same factors influence politicians all the time. To some extent of course any general theory which tries to cover events in a number of

countries over a considerable time period must provide an invariant framework to work within. But there are degrees of invariance, and it seems unlikely that politicians would always react in the same way to the wildly fluctuating vagaries of domestic and international politics, nor that ideological distances or policy agreements always remain the same. It would be desirable to have a theory which made some allowance for changing historical circumstances and politicians' reactions to them (Laver, 1974).

A deliberately simplified formal theory cannot, however, be adequately criticized on the grounds that politicians do not consciously apply it. They may, after all, act in the way it describes even if, for prudential or electoral reasons, they do not admit this or even are themselves not aware of doing so. The proof then rests on the extent to which actual coalitions form in ways specified in the theory. Unfortunately, when we look at the evidence, only 34% of coalitions in 12 Western European countries between 1945 and 1971 are covered by the optimal combination of minimal winning and ideological distance criteria (Herman and Pope, 1973). An additional 30.2% were formed by surplus majority governments, i.e. by a coalition of more than enough parties to form a minimum winning group. And another 35.8% were composed of coalitions with less than 50% of legislative seats – an even queerer phenomenon if opposition politicians were indeed ruthless office-grabbers.

These conclusions hold when governments are defined as administrations formed after a general election and continuing in the absence of:
(a) change in the Prime Minister;
(b) change in the party composition of the cabinet;
(c) resignation in an inter-election period followed by reformation of the government with the same Prime Minister and party composition (Hurwitz, 1971; Sanders and Herman, 1977).

This is a standard definition and it is the one adopted throughout this discussion. Normally we do think of such events as defining a government's period of office, and it brings our theories closer to practice when we use the common-sense definition. In a study which omits the criteria of elections and resignations, and thus assimilates governments with the same Prime Minister and party composition, Laurence Dodd (1976) does establish a tendency for governments formed on minimal winning principles to last longer. The difficulty is to sort out the implications of this when, for example, the 1949 Menzies Government in Australia lasted by this criteria for 17 years! There are, after all, reasons why governments resign, even if they re-form later along the same lines. It is likely, after this crisis, that power relationships and/or policy priorities have changed, so it makes sense to distinguish the two.

Recent work by Schofield (1983, Chapters 2 and 3) using the same definition of government change as Dodd, has confirmed the tendency for

minimal winning coalitions to last longer. It further appears that minimal winning coalitions tend to form as expected where there are a reasonably small number of significant parties. As fragmentation increases, in the sense of the number of significant parties going up, surplus coalitions alternate in increasing numbers with minority governments. This could be interpreted as a reaction to the confusion and uncertainty inherent in dealing with too many independent actors, and consequent difficulties of calculation. In turn the strain of making constant concessions and compromises seems to provoke internal splits and disagreements within the existing parties, thus increasing fragmentation in a continuing vicious circle.

These findings indicate that in certain circumstances the office-seeking, minimal-winning criteria are relevant. However, the criteria are not universally applicable and indeed may apply only under rather specific circumstances. This point is reinforced by the findings of Taylor and Laver (1973) and Hermand and Pope (1973) using the generally accepted definition of government given above, that roughly a third of West European government coalitions did conform to minimum winning criteria (tempered, however, by considerations of policy distance). The new formulation outlined below is compatible with these findings in allowing for the application of minimal winning office-seeking considerations where policy agreement is lacking (see Table 2.1 below).

Because minimal winning criteria apply only patchily, steps have been taken to formulate rational choice criteria for coalition formation, based on policy distance of some kind which almost, or even wholly, exclude office-seeking from consideration (De Swaan, 1973, p. 156; Browne and Rice, 1979; Grofman, 1980). Such formulations may quite happily predict the formation of surplus coalitions where the parties are ideologically contiguous. They have not yet been tested against a wide range of systematically collected evidence. But the fact that large numbers of surplus majority governments are known to form, attests to their realism in this respect. This shift from office-seeking to pure ideological or policy considerations in coalition theory suggests the need to give explicit precedence to policy, as in the general theory of government below (Section 2).

The evaluation of office-seeking theory has concentrated on the emergence of government coalitions because it is here that it is most explicit and widely applied. Evidence on the allocation of ministries is ambiguous. Browne and Franklin (1973) related the share of ministries received by parties to their share of seats, and noted a strong proportionality between them. Parties, in other words, received an allocation of ministries proportional to the share of seats they had contributed to support the government. While this might seem to fit a straightforward office-seeking interpretation, maximizers of office among smaller parties crucial to the formation of government would surely demand a disproportionate share

of ministries (cf. Olson, 1966). Since in absolute terms this would still be worthwhile for larger parties to concede (for relatively few offices are involved relative to those which can be gained), one would expect on strict office-maximizing assumptions to find strong disproportionality in favour of small parties. Some tendencies in this direction are noted by Browne and Franklin, but not to the extent that would be envisaged by the theory.

Strictly speaking, office-maximization has been applied only to coalition formation and to some extent to the allocation of ministries, rather than to other aspects of government activity. As noted above, its close relationship to the office-seeking theory of party competition nevertheless implies that politicians will try to carry through policies supported by the majority of electors, since that is what will ultimately secure victory and office. Since this is the only consideration in their minds, past commitments or long-standing ideological attachments will carry no weight compared to current majority preferences (cf. Budge and Farlie, 1977, Chapters 4 and 5).

On the other hand, the postulate of rapid policy adjustments of this kind contradicts the insertion into office-seeking coalition models of fixed ideological positions which render cooperation between neighbours more profitable than cooperation across greater ideological distances. If positions are constantly being adjusted to meet electoral preferences, fixed ideological positions will not exist, even for limited periods of time. Calculating transitory proximities will rapidly become unmanageable.

Thus the promised integration of the two branches of office-seeking theories, which in the abstract forms an attractive explanation of most electoral and governmental processes, actually conceals an important anomaly in regard to the extent and speed of policy adjustments. Either the office-seeking coalition theory has to drop assumptions about stable ideological positions for the parties, which would diminish its fit to existing evidence, or office-seeking theories of electoral competition have to modify their assumption that parties are infinitely flexible in policy terms.

In point of fact parties do seem to have stable policy commitments, to which in some cases they subordinate office but which in any case are bound up with their political aspirations. Because of the impossibility of making fine adjustments and calculations, electors vote on the basis of fixed associations between the parties and certain broad policies (Robertson, 1976; Budge and Farlie, 1983, Chapter 2). Certainly party strategists act on this assumption (Budge and Robertson, 1984). So parties have no hopes of evading the connection. If they do not carry through their characteristic policies while in government, they will not profit by their neglect (quite the reverse in some cases).

This interpretation saves the postulate of relatively fixed ideological positions for the parties, which has been incorporated in some form into most office-seeking models. But it goes further in pointing to the domi-

nance of policy considerations over a strategy of immediate office-seeking. Alterations to the original models have themselves mostly followed this direction, to the extent that policy considerations have replaced office-seeking altogether, as noted above.

Given the poor overall fit between office-seeking models and the actual behaviour of parties in government, the obvious way to build a new and more satisfactory theory is to turn the original formulation on its head, and give primacy to policy considerations. Instead of assuming politicians to aim at personal aggrandizement, we postulate that they put preferred policies above office. Participation in government is valued because it gives parties an opportunity to implement policy. Such an assumption links up party behaviour in government to what has been shown to happen in elections (Budge and Farlie, 1983). It also extends the motives behind the formation of government (Chapter 2) to those determining the distribution of ministries (Chapter 3), to the principles of government policy-making (Chapter 4) and to the purpose behind ministerial reshuffles (Chapter 5). Although this theory of democratic government breaks with earlier theory on the primacy of policy over office, the break is not complete. The hierarchical principles postulated to underlie government formation leave a place for minimal winning coalitions in multi-party systems without policy agreement (Table 1.2). Prime Ministers who push preferred policies act indistinguishably from Prime Ministers who wish to keep themselves in office (Table 1.3). Formally, of course, the statement of theory is inspired by the shape of the earlier explanations. More broadly, the pursuit of policy seems quite as "rational" as the pursuit of office, besides giving a more attractive picture of the democratic politician as an upholder of principle rather than a pursuer of place.

## 2. A comprehensive theory of democratic party government giving primacy to policy preferences

### (i) *Assumptions*

This section systematically develops the consequences of the critique in Section 1 and the tendency of office-seeking models themselves to incorporate ideology and policy. If policy preferences are taken as the major determinant of politicians' action, one must specify what these are and how they affect behaviour within governments. As will be seen, this still leaves a place for the minimal winning coalition, but as a last resort rather than as the general norm. As well as the shape taken by policy preferences, it must detail the circumstances under which a government can actually form (Does it require a minimum of 50% plus one seats in the legislature as the office-seeking argument implies? Or can it make do with less now

*Table 1.1*
*General assumptions of an integrated theory of democratic party government*

1. In parliamentary democracies the party or combination of parties which gains a majority on legislative votes of confidence forms the government.
2. Parties seek to form that government with a majority on legislative votes of confidence which will most effectively carry through their declared policy preferences under existing conditions.
3. (a) The chief preference of all democratic parties is to counter threats to the democratic system.
   (b) Where no such threats exist, and Socialist-Bourgeois differences separate the parties over salient current issues, the preference of all parties is to carry through policies related to these differences.
   (c) Where neither of the preceding conditions hold, parties pursue their own group-related preferences.
4. Normal governmental arrangements are most effective in getting policies carried through. Subject to their declared policies being advanced, therefore, parties seek to form governments with a party composition as close to the normal as possible.
5. Within parties, and subject to overall policy agreements and disciplinary and procedural constraints, factions seek to transform their own policy preferences into government policy most effectively.
6. With the exception of essentially caretaker administrations, government ministers, including the Prime Minister, are members of parties; and within them, of factions.
7. Ministers are replaced if forced to resign from their particular post.

the assumption of pure office-seeking has been modified?). Parties are not necessarily united internally, of course, although their endorsement of a large number of common interests renders it natural for many purposes to view them as single actors. This is entirely defensible for purposes of simplification, when (for example in the formation of governments) most parties have clearly defined purposes and positions on which most of their members at least temporarily agree. In dealing with the functioning of governments over a one to four-year period, however, internal party factions emerge, even if they were not there before (but they normally are). Their existence affects the policies governments make, and perhaps even more closely the reallocation of ministries and reshuffles which occur during the lifetime of most governments.

Table 1.1 spells out the general assumptions of a new theory designed to cover these and related points, and to provide an integrated basis for explaining all the major aspects of party behaviour in government.

The assumptions are stated verbally, as precisely and clearly as possible. The advantage of presenting them together in a table is to render them more easily memorable and to emphasize their inter-relationships. A tabular presentation serves also to highlight any ambiguities or unforeseen complexities, which the object is to eliminate. In the table the assumptions

form a progression from those necessary to cover the question of government formation to those dealing with other aspects of government (though most assumptions have implications for more than one area, as will appear).

Although it might be regarded as purely definitional in nature, Assumption 1 does specify the minimal condition for a government's existence – the state of affairs which parties must aim to create if they wish to form a government and which they must maintain if they want to stay in office. At first sight it may seem obvious (in the parliamentary regimes being discussed) that governments need a majority in legislative votes of confidence. The ability of the government to gain legislative support is, after all, specified in most constitutions (written or unwritten) as a legal requirement for its survival.

Note, however, what this very bare assumption is *not* saying – that is its real significance in our argument. It does *not* maintain as an office-seeking approach does, that the government must have a majority of seats in the Chamber in order to win votes of confidence, on the further assumption that all non-government parties impelled by the urge for immediate office will gang up on the government at every opportunity. Since office-seeking has been abandoned as the primary motivation, there is no reason why opposition parties may not offer voting support to a government from outside, or at least abstain, if they have policy aims which may be served by these courses of action. Moreover, two non-government parties, even if they dislike the government and wish to replace it by themselves, may still like that alternative better than its replacement by the other. Hence they will rarely vote simultaneously against it. For all these reasons a government may survive quite comfortably for a considerable period with less than 50% of legislative seats – indeed, quite often with less than 40%. Thus Assumption 1 is quite compatible with the observed frequency of minority governments – which are almost as common as minimal winning coalitions.

With the question of what is at stake clarified, Assumption 2 states explicitly the point that has been made generally above – that parties' main concern is with setting up governments which will carry through their preferred policies, rather than simply being in government for its own sake.

Assumption 3 states explicitly what parties' preferred policies are. In order of priority these are to conserve democracy, if that is in any sense under threat; where such a threat is absent to deal with distributional and planning matters related to Socialist policies (whether the parties concerned are for or against them); and where these are not salient, to pursue their own characteristic policies (which are most often designed to benefit the social groups from which the party draws its support).

In making such explicit assumptions about parties' and politicians'

preferences, the new approach breaks decisively with earlier ones. Under the desire to preserve generality wherever possible, most formal models seek to make their assumptions and implications compatible with any general aim which actors may hold. On the other hand, since it is necessary to postulate certain goals for actors in order to give the formal reasoning some application and content, this usually involves a narrowing down of postulated goals to pure office-seeking, on the ground that attainment of office is prerequisite to pursuing any policy. While this reasoning has a surface plausibility it is obvious that it breaks down in the many cases where the price of entry to office is the abandonment or even drastic modification of preferred policy.

Politicians, on the contrary, usually seem to have certain specific policies in mind when they set up or enter governments. These policies vary with circumstances. Politicians, like anyone else, tend to switch their priorities in response to changed situations. The wording of Assumption 3 allows for this. Thus in a crisis of the regime, it is highly unlikely that the very politicians who have been most influential in shaping its practices, and who in many ways are the chief beneficiaries of existing arrangements, will not make their main concern its support and defence. Such politicians will, by definition, be in the majority since if they were not, the democratic regime would already have ceased to exist! Whatever their previous differences over social and economic policy, or in regard to the various group interests they represent, they will come together when the system of bargaining whereby these differences are settled is itself attacked. In that, they have a common interest overlying the disputes they carry on within the bargaining process.

In a normal political situation, democracy is not seriously threatened. This gives free play for such "ordinary" disputes to emerge. The most pervasive and general of these relate to the programmes of Socialist parties or related "progressive" parties (Budge and Robertson (ed.), 1984, Chapter 20). Whether or not these are full-blooded Marxist or simply reformist in intent, they all threaten upsets in established relationships which tend to involve most people in the society. As a result, all parties will tend to line up on either the Socialist or anti-Socialist side when these issues come to the fore, sinking the differences which at other times might have separated them. Under such circumstances they will be willing to refrain from pushing points to which at other times they should have been firmly wedded, since their clients would, in their view, be more decisively affected by the success or reversal of the Socialist programme where there seems a real possibility of it being effected.

Where this is unlikely, and no threat to the democratic system exists, then parties and politicians are free to pursue the policies most characteristically associated with them. As we have noted, these are often interests or needs of supporting or associated groups. Socialist parties will follow this

course too, but in a more incremental and piecemeal fashion than in periods of outright confrontation with bourgeois parties.

Assumption 3 is, of course, specifically phrased so as to cover the post-war situation in the democracies covered here. It is probable that the same specification would apply pre-war, although its applicability to that period will not be checked here. Although the assumption does tie democratic politicians down to a set of substantive preferences, it is hard to see what other motives could be operating in their situation, whether in the post-war period or at another time. So the loss of generality may be more inhibiting for mathematical manipulation than for substantive explanation.

Assumption 4 might almost be termed the principle of inertia operating within our framework. Parties are concerned with setting up the government arrangements which will most effectively attain their policy preferences. To change the normal arrangements more drastically than is required for that purpose is self-defeating, as this might provoke all sorts of alarms and resistance, thus rendering it harder to carry the preferred policy through. This is because overturning established arrangements is both a signal that further, more drastic changes are to be expected (thus giving opposition more time to rally) and anxiety-provoking in itself. In a sense, usual political guarantees are being withdrawn. Thus there is a premium – even for parties bent on radical change – on associating with themselves some party which has often been in government.

This consideration applies mainly, of course, to parties in a coalition system, where governments tend to be made up of two or more parties. But it is not absent from two-party majority systems, where one party constitutes a government on its own. Normal governments then consist of one or the other of the major parties. Where one of these parties is being replaced, however (as in Britain the Liberals were by Labour in the 1920s), it will always be prudent for the newer party to associate itself initially in government with one of the older parties.

We have commented on the postulate of party unity and common purposes. Assumptions 2 to 4 in fact ascribe common preferences and motivations to parties as such. This is realistic in that any party which keeps together as a functioning entity must preserve a minimal set of common purposes *vis-à-vis* other parties. Office is rarely enough to provide a unifying bond, because as already pointed out the mere achievement of office gives little satisfaction if it does not provide a means of cooperating with like-minded individuals to effect policies. If a party is so divided as not to agree internally on some preferred policies more than its factions do with groups outside the party, then no motivations exist for keeping it together.

Within this set of common purposes there is, however, room for differences of emphasis and priority. Within most parties these give rise to

different groups and factions, which compete to place their own adherents within ministries and to influence government policy. Assumption 5 explicitly recognizes the existence of such factions and groups, but also incorporates the qualifications made above – their struggles are contained within the overall party framework. This consists both of shared, overriding, policy agreements and sanctions which will be brought to bear upon those who violate these.

Most individuals in a party will belong to a faction, and ministers and Prime Ministers are no exception. Their loyalty to their faction as well as to their party will motivate their behaviour, with consequences which will be most evident in the area of internal government change. Specifically related to this area is Assumption 7, which makes the perhaps trivial but none the less essential point that ministries do not normally disappear with ministers, so the quitting of office by an individual for whatever reason provides the stimulus for a change in the composition of the government. We shall follow through the consequences of this in relation to Table 1.3 below.

(ii) *Implications for government formation*

Since governments have to be formed before they can be reshuffled, we first, however, deal with the implications of our assumptions (specifically Assumptions 1 to 4 in Table 1.1) for the way in which governments are initially constituted. Here we are primarily concerned with which parties enter the government, rather than with the question of how ministries are shared between them. We deal with this later (Section 2 (iv) below). As in the case of office-seeking theory, the implications from the assumptions take the form of criteria specifying what the party composition of governments will be under given distributions of legislative votes between parties. They are stated in Table 1.2. Criteria i, ii(a), iii and iv derive from the first three assumptions of Table 1.1 and Critera ii(b) and v from Assumptions 1 to 4 of that table.

The relationship between criteria and assumptions is clear-cut. If (by Assumption 3(al politicians' chief concern when the democratic regime is threatened is to defend it, it must follow that all pro-system parties will seek the most effective means of doing so. This is to form a government of national unity, far in excess of the bare numbers needed to survive votes of confidence (Assumption 1). Only by staging an unusual show of unity and determination can threats (whether external or internal) be outfaced. Such a "surplus majority" government is inexplicable in terms of pure office-seeking, but it is very understandable in terms of a general party agreement on a burningly important question of the day – a question, almost, of survival.

The desire to safeguard the regime will also incline democratic parties

*Table 1.2*
*Hierarchical criteria for government formation implied by Assumptions 1-4 of Table 1.1*

| | |
|---|---|
| Criterion i | Where the democratic system is immediately threatened (externally or internally) all significant pro-system parties will join the government excluding anti-system parties. |
| | *In the absence of immediate threats to democracy:* |
| Criterion ii(a) | any party with an absolute majority of legislative votes will form a single-party government; |
| Criterion ii(b) | except where such majorities are unusual, where it will form the dominant party of a government excluding anti-system parties. |
| | *Where no party has a majority of votes and Socialist-Bourgeois differences over current issues are salient:* |
| Criterion iii | the tendance with the majority will form a government either including or with support from all numerically significant parties in the tendance (anti-system parties can only provide support and are excluded from participating in government). |
| | *If no such Socialist-Bourgeois differences exist the party which:* |
| Criterion iv(a) | is largest and has a near-majority of votes; |
| | or |
| Criterion iv(b) | is manifestly larger than any other pro-system party will either form the government alone in countries where single-party government is normal, or will form the dominant part of a government (excluding anti-system parties). |
| | *Where Socialist-Bourgeois differences are not salient and no single party has sufficient votes to meet Criterion ii or iv, coalitions with a plurality will be formed:* |
| Criterion v(a) | to group the parties most agreed on the specific issues currently salient; |
| Criterion v(b) | failing such agreement to minimize the numbers of parties in government to those which will provide a majority on legislative votes of confidence; |
| Criterion v(c) | in any case, to include the normal parties of government (if any) subject to v(a) and (b), and to exclude anti-system parties. |

to shun cooperation with parties opposed to it, under any circumstances. This is not an implication of the assumptions which appears as a separate criterion on Table 1.2, but it does appear throughout as a constraint on the composition of any coalition government, within the other criteria.

Where no threat to the regime is perceived, Criterion ii in Table 1.2

states that any party with an absolute majority of legislative votes will form a government on its own, or (with an eye to the normal arrangements in a country where absolute majorities are unusual) will at least dominate the government, probably in association with a small "party of government" (i.e. a minor party which almost always participates in the ruling coalition). Criterion ii(a), on single-party government based on an absolute majority of legislative votes, of course covers the classic situation produced by competition between two evenly balanced major parties. Whether differences over the Socialist programme are salient, or in their absence the party is concentrating on its own characteristic, group-related policies, single-party government will form the easiest way to achieve its preferred goals. The same applies, of course, where for prudential reasons a small party has shared in government even though this was unnecessary in purely numeric terms.

However, taking democracies as a whole, the emergence of an absolute majority for one party is relatively rare. In its absence, a quasi two-party system may temporarily be created by a resurgence of Socialist/non-Socialist divisions. As we have suggested, the repercussions of a full-blooded Socialist programme affect most people in the society quite strongly. Support or opposition to these forms a cement between parties which overrides normal conflict. The salience of the Socialist programme thus creates a quasi two-party competition between opposing ideological tendencies, in which the opposed coalitions act like majority parties. In such situations there is no question of some of the parties on one side being detached to cooperate in government with the other. Since the Socialist/non-Socialist cleavage is central to party competition, such an event would be as unthinkable as some faction of a majority party joining the opposition. Even anti-system parties excluded from actual government participation are driven by ideological imperatives to offer support to their own side in this cleavage.

Socialist or non-Socialist loyalties, where these become salient, are thus perfectly capable of providing a strong basis of support for government, even in a comparatively fragmented multi-party system. Their saliency, like the existence of anti-democratic threats, will, of course, vary over time. In some countries the prevalence and importance of division within the Socialist and non-Socialist camps themselves may prevent the differences separating them from ever coming to the fore of politics. In such cases, and also at times when these divisions are muted by other events or the passage of time, other arrangements emerge.

We are talking, it will be remembered, of situations where no party has a majority of legislative seats (otherwise, by Criterion ii, it would automatically form and dominate a government). However, there are a wide variety of intermediate situations between the emergence of a single majority party and complete fragmentation into a range of small parties.

The case considered next is a dominant party system with one outstanding party – outstanding either because it has nearly missed a majority or because it is obviously larger and more important than any other legitimist party. Since a democratic regime is unlikely to continue functioning where an anti-system party is largest, this last characteristic also implies that it is outstanding in relation to all the other parties.

Such a dominant party is likely to have been repeatedly in government in the past, so its claims to office are enhanced by the desirability of preserving normal arrangements. Its size renders it the obvious basis for building an administration – a consideration reinforced by its ability to bring down most governments excluding it. Such a party can well, with the tolerance of non-government parties, form a viable government on its own. Failing tolerance, it may be able to rely on divisions between the other parties keeping it in power as the most acceptable alternative.

Of course, the position of a dominant party is strengthened if other parties will join it in a governmental coalition. Its contribution assures it a directing role in the government anyway, and the continuance of the government is more assured through the adherence of other parties.

The absence of a majority party does not, therefore, preclude the possibility of government formation on the initiative of the largest party. Obviously the position of the government is weaker where it commands only a minority of seats, or is subject to the possibility of intra-party disputes. But it is not by any means untenable.

The absence of a dominant party serves to convert the situation into the classic case of multi-partism, where a considerable number of relatively equal parties exist – none being of sufficient weight numerically to give it an outstanding role. In this case none of the parties forms such an obvious basis for the government as to force the other parties to negotiate with it. Coalition-building must proceed instead by negotiation between equals.

Threats to the regime or Socialist-bourgeois tensions will not provide a common focus for the negotiations since by definition they do not exist in the situation we are discussing (if they did, either a surplus-majority government would form under Criterion i, or a quasi-majority tendance government under Criterion iii). There is still a possibility of attitudinal cement being provided for a government through agreement, however, since in the absence of other concerns each party will have its own characteristic policies which it wishes to pursue. If potential partners' concerns do not conflict too much on the questions which are currently salient to them, a common programme can be hammered out as a basis on which to form a government. Such agreement is important since it reduces tensions and costs of internal negotiation. It renders joint progress towards desired goals more likely, thus increasing satisfaction with the existing alliance and averting the possibility of the government foundering amid mutual recriminations.

The negotiation of such a programme is one way of proceeding in a fragmented multi-party system. It is entirely possible, however, that party preferences conflict so much on salient current issues that no genuine agreements can be negotiated. Governments, on the other hand, must still be formed. Since the bargaining process cannot be based on ideology or policy, the pursuit of office becomes paramount. Here we enter the situation postulated by minimal winning theories – but only because prior policy-based considerations are, by definition, non-operable. In default of policy agreements, it is certainly advisable to keep the number of parties in the governing coalition as small as possible, to minimize the costs of disagreement and internal negotiations. The size will be set, however, not by the need to gain an absolute majority of all legislative seats but by the need to group parties with enough support to win legislative votes of confidence. Policy *agreements* may be absent but enough *disagreements* can exist among the major non-governmental parties to prevent them all lining up against the governing coalition at one time. We should consequently expect parties in such a situation to form a combination of the least number necessary to attain a consistent majority on votes of confidence. (Operationally, as explained in Chapter 2, we should know what level of support is necessary for votes of confidence by taking the average number of seats governments have held in the past.)

Both the formation of governments by agreement and on the basis of a minimum winning combination are constrained by the other factors emphasized in Criterion v(c) – the need to include normal parties of government and to exclude anti-system parties. After our earlier discussion the latter requires no additional justification. The inclusion of normal government parties is of even greater importance to the fragile governments emerging under Criterion v, since the repercussions of not including them might well be enough to upset the whole arrangement. Besides, such parties are often the ones motivated to begin negotiations with other parties and to smooth over differences between their partners. Their familiarity with administration also removes another potential point of friction from an already trouble-prone situation.

Even where policy agreements are absent, the minimal winning coalition may not be quite as small as in abstract it might be, owing to the need to include established government parties. We shall take this factor into account when reviewing the success of the criterion in Chapter 2.

The important point to note is the incorporation of the minimal winning criterion (realistically modified) in our "rules" of government formation. This takes cognizance of the findings cited in Section 1 on the incidence and longer duration of such governments in certain situations. As the earlier critique stresses, however, the criterion becomes relevant only when ideological and policy considerations do not provide an alternative basis for the creation of a government. It is a procedure of last resort,

applied in the absence of other, more satisfactory, modes of action. While our formulation offers a synthesis with earlier theory, it does so on an assumed primacy of policy considerations. This also carries over to the other implications considered below.

(iii) *Outside support of governments*

On pure office-seeking criteria, parties which consistently or generally support a government of which they are not part are acting with total irrationality. Since they do not share in ministries or other spoils, they can have no comprehensible reason for offering support, apart from building bridges to the next government. But in office-seeking theories parties have a short time perspective – as is only realistic in politics. If we view the matter from a concern with policy, however, there are very good reasons for some excluded parties supporting governments. This is whenever the existing government, however unattractive, offers a better chance of the party's policies being put into effect than any of the other likely alternatives. Even where the government's policy is only marginally better from the viewpoint of the party, the costs and trouble involved in a governmental crisis may well induce it to maintain support at crucial junctures.

This point is spelled out in Implication 1(iii) of Table 1.3 (which summarizes other applications of the assumptions). It is a difficult proposition to test since it carries the danger of circularity – immediately an outside party supports a government this might be taken as proof that it must see the alternatives as worse. However the implication need not be circular provided there is direct evidence on how party leaders view the situation. And it does fit the real situation in which such outside support is often given, where the office-seeking hypothesis does not. Chapter 2 deals with some of the operational difficulties of testing this point.

A particular case of parties offering support without participation comes with anti-system parties. These are generally excluded through the deliberate reactions of the other parties, inspired by their overriding urge to preserve democratic procedures (Implication 1(i]. For anti-system parties themselves, support will be most clearly called for at times when the Socialist programme is salient, as it strongly affects their own clients and policies. In such a situation they will line up with the other parties on their own side of that cleavage, in order to maintain a government of their particular *tendance* (Implication 1(ii)).

(iv) *Distribution of ministries*

Most theories of democratic government confine themselves to the party composition of coalitions (taking single-party governments as the extreme

*Table 1.3*
*Major implications of the general assumptions of Table 1.1 for other parties' support of government from outside, for the distribution of ministries, for policy-making, and for dissolution of governments*

1. *Support of government rather than participation*

(i) A party regarded as anti-system cannot because of the opposition of other parties participate in government, and can only support and not join governments which will pursue some of its preferred policies (Assumptions 1-4).

(ii) Where threats to democray are absent but Socialist-Bourgeois differences are salient, anti-system parties will support parties of their own tendance from outside government (Assumptions 1-3).

(iii) A party which cannot persuade others to form a government which will put into effect any of its characteristic policies, will not join the government which is formed, but will vote for/abstain in favour of that government if it considers all practicable outcomes (including an election) would further reduce the possibility of putting its policies into effect (Assumptions 1-3).

2. *Distribution of government ministries between parties in a coalition*

(i) The largest party in a coalition will take the premiership (Assumptions 1, 2, 3(c].

(ii) Subject to rough overall proportionality, each party will seek control of ministries in their own areas of policy concern, e.g. Agrarian parties will seek the Ministries of Agriculture and Fisheries and Regional Affairs; Labour parties will seek Ministries of Social Affairs, Economic Affairs, Labour Relations; Conservative parties will seek Defence, Interior, Justice, Foreign Affairs, etc. (Assumptions 2, 3(c].

(iii) Where a particular type of party does not exist, the most similar of the existing parties will seek ministries in its area of policy concern (Assumptions 2, 3(c].

(iv) These tendencies are least evident when governments are formed to counter anti-democratic threats and less evident when *tendance* governments are formed in a situation of Socialist-Bourgeois hostility.

(v) A small party in a government which could be formed by a large party on its own will not necessarily get a proportionate share of ministries.

3. *Policies pursued by governments*

(i) Where there is a threat to the democratic order, governments will direct their main policies to countering it (Assumptions 1, 2, 3(a].

(ii) Where there is no such threat but Socialist-Bourgeois differences do separate the parties over salient current issues, governments will be chiefly concerned with redistributive policies and government control and intervention (the direction of policy being decided by the Socialist/Bourgeois complexion of the government) (Assumptions 1, 2, 3(b].

(iii) Each party in the government will have some of its preferred policies put into effect, i.e. governments including an Agrarian party will pursue policies more favourable to farmers and rural interest than governments without an Agrarian party: similarly with Labour parties and the working class and Conservative/Liberal parties in regard to business, etc. (Assumptions 2, 3(c].

4. *Turnover of personnel*

(i) The death/illness/withdrawal of a member of a government always produces a reshuffle (usually limited) (Assumption 7).

(ii) The turnover of individuals in ministries is greater where the Prime Minister has more freedom of action; and declines as the Prime Minister has less freedom of action in relation to:
(a) other ministries,
(b) party factions,
(c) coalition parties in government
(i.e. where the Prime Minister has greater opportunities he is able to move factional opponents to prevent their creation of a power base, to take action in event of failure by a minister, to conserve general prestige of government, etc.) (Assumptions 1, 2, 3(c), 5).

5. *Dissolution of governments*

(i) When a threat to the democratic order ceases, governments formed to meet it terminate.

(ii) When Socialist-Bourgeois differences cease to be salient, *tendance* governments formed in relation to them terminate.

Where there is no threat to the democratic order and Socialist-Bourgeois differences are not salient:

(iii) When a Prime Minister can fix the date of an election, he will dissolve government when he feels confident he can improve his party's vote share in an election (associated with a good government record) or lessen vote losses (Assumptions 1-3).
(iv) When the Prime Minister is defeated within the government on a major current policy by another party/party faction, he will resign and effectively dissolve the government (Assumptions 2-6).

(v) hen disagreement on major current policies or ministerial replacement provokes withdrawal of support by a coalition party partner, the government will dissolve (Assumptions 1-3, 5-6).

(vi) Where in a coalition government one party has markedly declined in popularity, it will withdraw and force a dissolution of government (Assumptions 1-3).

case of a minimal winning coalition). On office-seeking criteria, however, the distribution of ministries is equally or even more important, since control of a ministry is the main reward sought by parties for entering government in the first place. Even on policy criteria, control of ministries is important, because it is the crucial element in the formulation and (perhaps more important) the implementation of programmes in a particular area. All this is simplified in the case of a single-party government, of course, since all ministries are then at its disposal and the problem of distribution then relates to intra-party factions – with which we deal below (Section 2 (vi)). Most of the discussion in this section relates to coalition governments, normally of several parties.

The few treatments which have dealt explicitly with the distribution of ministries, mostly on an office-seeking basis, have stressed proportionality as the major criterion for allocating offices within a coalition. That is, each party expects, and gets, a share of ministries more or less equal to the share of legislative votes it contributes to government support. Empirical evidence for this equivalence has been produced from a matching of seats and ministries for European cabinet coalitions from 1945 to 1970. This produced an almost one to one equivalence between shares of seats and shares of ministries (Browne and Franklin, 1973). The authors of this study noted that there may also be a form of "qualitative" as well as "quantitative" proportionality in operation, whereby the allocation of "important" ministries may supplement the quantitative norm. The only available evidence on this aspect of the distribution – for coalitions in Indian State Governments – indicates that larger parties did not necessarily receive the more important ministries (Bueno de Mesquita, 1975).

A policy-based approach must put more emphasis than the office-seeking formulation on parties' concern not just with important ministries, but with ministries in the particular areas of their interest. Thus it is not just a question of getting an equivalent return for their support in general, but of securing a specific ministry or ministries because of their significance for the party's policy concerns.

At the same time proportionality between seats and votes cannot be ignored. Precisely because all coalition partners wish to advance their own goals so far as possible, they will seek control of as many government ministries as they can and thus limit the numbers available to other parties. Each party has a sanction in that withdrawal will lessen the government's chances of survival. The most likely division that they can agree on will be one which (a) secures each party the ministry(ies) important to it (subject to the constraint that some of these may be of equal importance to other parties in the government, so not all ministries of concern may be secured); (b) maintains a rough equivalence between a party's support in the legislature and the number of Ministries it obtains.

This provision in regard to proportionality matches Browne and Frank-

lin's findings on European coalition governments. Whereas they interpreted this as evidence of a strict proportionality rule based on office-seeking, proportionality is regarded here as the second stage in a bargaining process directed at securing policy-relevant ministries. The crucial question in deciding between these interpretations is whether there is in fact any general connection between specific types of party and the types of ministry they obtain. This point is investigated in detail in Chapter 3. Implication 2(ii) in the table forms a summary of this argument.

It should, however, also be taken in conjunction with Implication 2(iii). Not all types of party exist in all countries, although the major types of ministry do. There is no Agrarian party in Britain, France or Germany, for example, and no Christian party in either of the first two countries. What happens in these cases to ministries which would otherwise have been allocated to the missing party? Its policy concerns are likely to be taken up by parties of a similar disposition, who seek to occupy the ministries it would otherwise have wanted to secure. For example, since the countryside is the place where traditional loyalties usually survive longest, one would expect religious parties to take up rural interests where a specifically agrarian party is absent. This point will be expanded in Chapter 3, through a specification of the policy interests of each type of party, and listing of the ministries that are as a result salient to them.

For larger parties with a developed comprehensive ideology, most ministries are relevant. Hence the amount and importance rather than the type of ministry is a prime concern. In this they will conform more to Browne and Franklin's rules of quantitative and qualitative proportionality. Not for the first time, office-seeking and policy-pushing criteria produce similar predictions here. The most important ministry in view of its centrality and dominance of the government agenda is the premiership. The largest party in the coalition can be expected to assert its claims to this, perhaps even being willing to cede otherwise salient ministries and to take somewhat less than its strictly proportional share in order to get it. The rationale here is that all parties covet this central post for its policy advantages: in resulting struggles or bargaining, the party with most resources will generally get it and this will be the largest party.

These are the major, general findings expected over all coalitions. There are the further Implications 2(iv) and 2(v), however, which state conditions under which this type of share-out will be less evident. This follows from the assumptions that coalition governments differ considerably according to whether they form in response to an anti-democratic threat, as appendages to a majority or dominant party, as quasi-majority coalitions of a Socialist or Bourgeois tendance, as agreed multi-party cabinets in the absence of the preceding conditions, or as minimal winning groups unable to agree on current issues. In the case of a government of national unity, where the overriding imperative is preservation of the

regime, it would be unreasonable to expect parties to stick out for ministries of their own concern. Somewhat similar considerations apply to participation in a tendance majority where the main consideration is the victory or defeat of the Socialist programme rather than parties' specific policy interests. Where a small party tags along with a majority or dominant party it can similarly not expect to get more than basic demands – perhaps the one ministry of most pressing concern (provided this is not also desired by the dominant partner). The greatest scope for bargaining comes with coalitions of many relatively equal partners. The general trends mentioned in Table 1.3 should be more evident here – and even more among relatively disagreed minimal winning coalitions than among coalitions formed on the basis of agreements on current policy. In the latter case, parties have through the coalition agreement some guarantee that preferred policies will be pushed regardless of the particular ministries they control. Where agreement has proved impossible, however, parties have no guarantee of getting their own way other than through control of particular ministries. Hence there should be an additional premium on occupying those of major concern to each partner.

### (v) *Policies pursued by governments*

All these points will be covered in detail in Chapter 3. The discussion now turns to related questions of policy-making. In a policy-based approach this is central: far from governments adjusting policy in order to gain office, its implementation is the major reason for parties taking office in the first place.

Since it is so central, the question has been thoroughly discussed already. So the way in which implications follow from assumptions is obvious. Governments will give first priority to support of democracy where it is endangered (3(i)). Where Socialist-Bourgeois tensions are high, governments will be mostly concerned with the points at issue between them, typically the distribution of income and other resources and extension of government intervention (3(ii)). And where neither consideration prevails, parties will be concerned with their own characteristic policies (3(iii)). We do not rely on the sketchy characterization of these made in Table 1.3 but spell out the connections between parties and their preferred policies, with supporting argument, in Chapter 4.

Generally the office-seeking explanation has neglected questions of policy-making. The only answer it has provided to the question of what governments actually do when in office derives from the related theory of party competition: parties in government adjust policies to the preferences of the majority of electors so as to gain more votes in the next election.

This is not a very clear or specific answer, however. There may be no clearly majority. Moreover, in a coalition each constituent party would work to please its own clientele since by definition in a multi-party system electors' opinions diverge widely. Where parties seek to enhance their reputation by acting reliably and responsibly this involves fulfilment of election pledges. Since these are the parties' declared policies the general office-seeking theory of party competition (but not its spatial formulation (cf. Budge and Farlie, 1977, Chapter 3-5] becomes equivalent at this point to the policy-making one: both concur on the assertion that parties in office will pursue their characteristic policies. Nevertheless the policy-making formulation arrives at this implication by a more direct route, and is to be preferred (if the implication is upheld in comparison with the evidence) on grounds of theoretical parsimony and clarity.

The other body of investigation which bears on government (as distinct from administrative) policy-making is that related to the "outputs" of government. This has generally concentrated on relating social and political characteristics of governmental units (class and occupational distribution, tax base, percentage votes for parties, etc.) to expenditures of government in various areas. The main finding from this line of research has been the influence of social and demographic factors as compared to the relative unimportance of party control or electoral strength. The main thrust of "output" reserach thus challenges the implications of the "policy-making" theory that party control makes a considerable difference to what governments do.

Recently, however, the thesis has been challenged for mixing such disparate governmental units at varying levels of development that socio-economic factors are bound to predominate. Where a country completely lacks resources, it cannot, obviously, finance a welfare programme. Where comparisons are based on units at broadly comparable levels of development, political differences – particularly the electoral strength of Right-wing parties – clearly show through. (For citations to the output literature and findings which show the overriding influence of party, see Castles (ed.), 1982.)

This finding is, of course, congenial to the thesis argued here. The various views will be contrasted with evidence in Chapter 4. Included among the countries under examination are governments like those of India and Sri Lanka which are at a lower level of development than the others. Nevertheless the analysis should escape the criticism of comparing unlike cases since it does not compare expenditure levels and other policies *between* countries with different party governments, but rather the performance of different types of government within the same country. Indian Governments in comparison with Western may be able to spend very little, but a Congress government should do different things from a non-Congress coalition.

(vi) *Turnover of personnel*

Not only does the office-seeking formulation neglect policy considerations, it completely ignores internal change within governments. Yet this is of great significance to their life and behaviour and any explanation seeking to be comprehensive must cover it.

The first implication (4(i)) follows directly from Assumption 7 in Table 1.1. Both incorporate the trivial but necessary point that the departure of a minister usually involves replacing him (rather than abolishing the ministry). He may be succeeded by a person from outside the current administration, but more usually by somebody from inside, who has to be replaced in turn. Thus the resignation of a single member usually produces repercussions which go beyond his particular post. Unless a major change had been impending, however, this type of event will not involve extensive transfers and turnover of personnel. Where affairs have been handled satisfactorily there is no inducement to disturb existing arrangements more than is absolutely necessary, and so the extent of change will be limited. On the other hand, given that senior politicians tend to be elderly, such enforced resignations are quite frequent and need to be explicitly noted as producing change extraneous to the main line of the other arguments.

These are taken up forcefully again in the second implication (4(ii)), which relates the extent of internal change and reshuffles to the power of the Prime Minister. We have already noted (Assumption 5, Table 1.1) that factions with distinct policy preferences will exist within most parties. Internally they will act in relation to each other just as externally parties do in relation to each other. That is, they will seek control of certain ministries within the overall party share, and try to advance their preferences by implementing them within ministries and seeking to influence overall party actions. Since factions will, by and large, agree more with other factions within the same party than with factions outside, overall unity will be preserved by an ability to negotiate compromises and by procedures for party unity and discipline which are explicitly designed to prevent disputes from getting out of hand. Nevertheless internal struggles, even though muted, may be expected to go on. Where there is a single-party government these will be the main source of government dissensions. In a coalition, of course, dissensions between parties overshadow internal factional jockeying and also put more of a premium on party unity.

By Assumption 6 in Table 1.1, we assume that the Prime Minister, like other members of the party, is a member of a faction committed to forwarding its policy emphases. He will advance these in part through his agenda-setting and related powers. To exert these he has, of course, to retain office and more immediately to prevent the emergence of alternative centres of initiative within the government. The most obvious way to

buttress his position and that of his faction is to move rivals fairly frequently, to prevent them consolidating a power base inside their own ministry.

Quite apart from helping his own faction, the Prime Minister has also to enhance the effectiveness and unity of the overall party so far as he can. This involves fairly prompt action to replace inefficient and unpopular ministers by better nominees.

These considerations apply mainly to single or predominant-party governments. In coalitions the Prime Minister's power is limited by the necessity of getting other partners' agreement to the replacement of their ministerial nominees. Unilateral attempts at replacement are liable to provoke a government crisis. Because of the difficulty of replacing ministers once a coalition agreement has been hammered out, internal change should be much less in the case of coalitions compared to single-party or predominant-party governments. Again, of course, one has to recognize the varying situations under which coalitions come into existence. Where an overriding sense of purpose binds the coalition together, as in the case of anti-democratic threats or Socialist-Bourgeois confrontation, partners are probably disposed to accept changes for the sake of maintaining unity. Where the coalition is simply a convenient tactical adaptation to the circumstances of the moment, change is more likely to result in crisis.

The incidence of reshuffles and replacements can then be related to restrictions on Prime Ministerial power, which are least in single and predominant-party governments and greatest in minimal winning coalitions without policy agreements. Factions within a party may, of course, be stronger or weaker, and to the extent that other factions are stronger the Prime Minister's freedom of action is less. It may also be constrained by institutional structures or constitutional conventions giving more autonomy to other ministries, although this is not likely to be the case in the parliamentary regimes with which we are concerned.

### (vii) *Dissolution of governments*

As noted, attempts by the Prime Minister to replace ministers in a coalition government may cause the breakdown of a coalition government, especially one composed of relatively equal partners and formed on the basis of ad hoc agreements, or even without these. This is recognized in Implication 5(v) of the table. Since each party has joined to advance its own policies as effectively as possible, a major threat to these – either in the shape of a direct attack or in the loosening of its control over a ministry – will cause withdrawal. This is also likely where a party sees its

future policy effectiveness impaired by continuing participation in the government (Implication 5(vi)). Both reactions are again more likely in tactical coalitions than in governments of national unity or in those based on Socialist or Bourgeous *tendances.*

In coalitions all parties hold an initiative in regard to the life of the government. In single-party governments such decisions rest mainly with the Prime Minister. As the representative of a particular policy line, he is likely to resign (or dissolve government and parliament if he has the power) if defeated on what he regards as a major point. If such a defeat occurs, the usual mechanisms for ensuring party discipline and unity have broken down and resignation forms the most attractive alternative. By this he may ensure an eventual triumphant return if he proves indispensable, or at least be able to mount a tough campaign for his alternative from outside government. These considerations also apply in a coalition government where the Prime Minister is acting to represent his party rather than an internal faction (Implication 5(ii].

In his role as party leader, the Prime Minister will also be concerned with long-term effectiveness and hence with electoral advantages. If he has the power of dissolution in his own hands, the point at which he chooses to exercise it will certainly be the most electorally advantageous for his own party when it stands a good chance of increasing its vote share (or attaining a majority of seats in competitive two-party systems). This should be at a point when the positive aspects of the party and government record outweigh the negative (Implication 5(iii)). Alternatively he may choose a time which will minimize his party's likely loss of votes.

Like all the other types of government behaviour discussed above termination and dissolution can thus be seen as reflecting party and factional concerns with policy. Where prospects for advancing declared preferences radically diminish, support for government continuance declines. Where continuance seems less fruitful for advancing preferences in the long run than dissolution, then the latter is adopted as a better strategy. This implies a bolder risk-taking approach to participation than would be implied by the office-seeking hypothesis, where selfish maximizers would be inclined to hold on to the last moment to enhance their gains from office.

All the implications discussed (5(iii) – 5(iv)) relate principally to periods when there is no overpowering reason for government unity. They are expected to apply where there is no threat to the regime nor confrontation of Socialist and Bourgeois *tendances.* In a modified form they may operate in the latter case. Additional reasons for government termination will, however, follow from the ending of the circumstances which have provided cement to hold a coalition together – the quietening of Socialist-Bourgeois tensions and the withdrawal of threats to the regime, for example. These are covered in Implications 5(i) and 5(ii).

## 3. Conclusions

The three summary tables provide a comprehensive, comparative theory of democratic government which integrates the various aspects of party behaviour at the level of politics. It is comprehensive because it can be applied to all the major aspects of governments' existence – how they form and change, what they do, how they end. It is comparative because it applies to all types of government formed by elected parties – single party, dominant party, or coalitions of various types. Hence it can be applied to all democracies, not just to Anglo-Saxon, Scandinavian, southern European or Third World. These are shown to differ because of specific divergences in the results of elections.

This theory is more extensive than existing formulations. It is also more plausible than office-seeking theories. Politicians may be out for themselves some of the time (and some politicians all the time!). But they also make stands of principle and have declared policies distinct from those of other parties to which they publicly commit themselves. If politicians merely wanted office, governments should be markedly stable once an initial division of offices was agreed. We know, however, that this is not so: they often fall apart over policy disagreements, while even within a single-party government factions emerge which are distinguished primarily on policies.

The policy commitment hypothesis also seems to fit existing evidence better than office-seeking, since it is compatible with all the types of government which actually form. The rest of the book is devoted to checking the fit further, by elaborating the implications of the explanation and checking these in detail against evidence from the 21 democracies. Chapter 2 examines the question of government formation; Chapter 3 the allocation of ministries; Chapter 4 policy-making; Chapter 5 internal change and termination. Chapter 6 presents an overall evaluation of results and discusses the particular modifications that they entail.

The development of this discussion follows the actual stages in which the project was conceived, starting from an a priori theory of government formation, extending its implications to other aspects of party behaviour in government, and then collecting and analysing information to check these. As the data were collected after developing the theory, its fit with subsequent findings should validate it more convincingly than is possible with a priori interpretation of previously available evidence.

## Bibliography

Axelrod, R., *Conflict of interest*, Markham, Chicago, 1970.

Browne, E.C. and Franklin, M.N., "Aspects of coalition payoffs in European parliamentary democracies", *American Political Science Review*, 1973, 67, pp. 453-469.

Browne, E.C. and Rice, P., "A bargaining theory of coalition formation", *British Journal of Political Science*,1979, 9, pp. 67-87.
Budge, I., Crewe, I. and Farlie, D.J., *Party identification and beyond: Representations of voting and party competition*, Wiley, London and New York, 1976.
Budge, I. and Farlie, D.J., *Voting and party competition: A theoretical critique and synthesis applied to surveys from ten democracies*, Wiley, London and New York, 1977.
Budge, I. and Farlie, D.J., *Explaining and predicting elections: Party strategies and issue outcomes in twenty-three democracies*, Allen & Unwin, London, 1983.
Budge, I., Farlie, D.J. and Laver, M., "Varieties of rational choice", Ms. Department of Government, University of Essex, 1982.
Budge, I. and Herman, V.M., "Coalitions and government formation: An empirically relevant theory", *British Journal of Political Science*, 1978, 8, pp. 454-477.
Budge, I. and Robertson, David (eds.), *Party strategies: An analysis of manifestoes and their equivalents in twenty democracies*, Sage, London, 1984.
Castles, Frank (ed.), Sage, London, 1982.
de Mesquita, B.B., *Strategy, risk and personality in coalition politics: The case of India*, Cambridge, New York, 1975.
De Swaan, A., *Coalition theories and cabinet formation*, Elsevier, Amsterdam, 1973.
Dodd, L.C., *Coalitions in parliamentary governments*, Princeton University Press, Princeton, 1976.
Downs, A., *An economic theory of democracy*, Harper, New York, 1957.
Gamson, W.A., "A theory of coalition formation", *American Sociological Review*, 1961, 26, pp. 373-382.
Grofman, B., "A dynamic model of proto-coalition formation in ideological N-space", Ms. School of Social Sciences, University of California, Irvine, 1980.
Herman, V.M. and Pope, J., "Minority governments in Western democracies", *British Journal of Political Science*, 1973, 3, pp. 191-212.
Hurwitz, L., "An index of democratic political stability", *Comparative Political Studies*, 1971, 4, pp. 41-68.
Laver, M., "Dynamic factors in government coalition formation", *European Journal of Political Research*, 1974, 2, pp. 259-270.
Leisersen, M., *Coalitions in politics*, Yale University, New Haven, 1966.
Olson, M., *The logic of collective action*, Harvard, Cambridge, Mass., 1966.
Rallings, C., "Comparing British and Canadian manifesto pledges with government action", Chapter 19 of Budge and Robertson (eds.), *Party strategies: An analysis of manifestoes and equivalents in twenty democracies*, Sage, London, 1984.
Riker, W.H., *The theory of political coalitions*, Yale University, New Haven, 1962.
Robertson, D., *A theory of party competition*,Wiley, London and New York, 1976.
Sanders, D. and Herman, V.M., "The stability and survival of governments in Western democracies", *Acta Politica*, 1977, 12, pp. 346-377.
Schofield, N., *Coalitions and European parliamentary systems*, Yale University, New Haven, 1983.
Taylor, M.J., "On the theory of government coalition formation", *British Journal of Political Science*, 1972, 2, pp. 361-372.
Taylor, M.J. and Laver, M., "Government coalitions in Western Europe", *European Journal of Political Research*, 1973, 1, pp. 205-248.

# Parties and Political Mobilization: An Initial Mapping*

HANS DAALDER
Leiden/Florence

## I. Introduction

This paper seeks to provide a tentative inventory of problems subsumed under the term political mobilization. It takes the use of the words "political mobilization" for granted, although certain queries have rightly been raised about the use of the term "mobilization". Thus one could note: the allegation that it has been uncritically transferred from the literature on totalitarianism to the study of open democratic political systems (Giovanni Sartori);[1] the objection that it is conceptually and grammatically mistaken to speak of a person mobilizing him1 or herself, instead of being mobilized by others (Gunnar Sjøblom); or the view that the term political mobilization should really be used only for the process of induction of new groups into a political system in the processes of mass democratization (as analysed above all by the late Stein Rokkan).[2]

In this paper we shall approach the concept of political mobilization in terms of linkage processes between different elite groups on the one hand, and citizens on the other. Action can be taken from either side. One of the major, and in a democratic era undoubtedly most legitimate, linkages is provided by what is usually termed the "partisan-electoral channel". In systems with responsible government, universal suffrage, and open, competitive elections, parties put up candidates and programmes in a competitive struggle for representation and office. This is held to ensure democratic accountability, and allows citizens to express definite preferences for persons and policies.

Obviously, parties do not provide the only, exclusive channels in the interaction between leaders and followers. There are a host of other intermediate actors, including the bureaucracy, organized interest groups, *ad hoc* action groups and the mass media. Both leaders and individual citizens can seek to short-cut party by more direct forms of contact, as in plebiscitary stances of leaders, or the seeking of direct access on the part of citizens. Moreover, neither citizens nor leaders come to their

interaction *tabula rasa*. Leaders inevitably have other links, i.e. with other parties, and the various other intermediate actors just mentioned. Holding offices also demands the fulfillment of specific roles and responsibilities. Citizens, on the other hand, are socialized into political life by numerous agencies beyond party, including the family, educational institutions, churches, recreational associations and the workplace; they are affected by experiences with alternative agencies like the bureaucracy or interest groups; and they also undergo life-cycle effects.

Yet for reasons of clarity it seems useful to proceed in the following analysis in three successive steps:

1. We shall first treat the role of parties as agencies of political mobilization *in abstracto*, assuming that one can consider the problem of linkages between parties and citizens in isolation from other intermediate agencies. Section III will explore problems raised in the relation between citizen and party (in the singular), Section IV in his relation to the party system (or parties in the plural).

2. We shall then consider the role of parties as agencies of mobilization, taking into account the interaction of parties with other possible agencies of political mobilization, including interest groups, action groups, the media and the bureaucracy. Given that this problem has very different aspects, whether one views it from the perspective of rulers, or from that of the citizens, the analysis will again be split in two sections: Section V traces the problems from the top downwards, and Section VI looks from the citizen upwards.

3. Finally, we shall present a few general propositions on trends towards a possible displacement of parties by other agencies of political mobilization.

But before we go into the analysis, some remarks will be made on historical mobilization processes (Section II). We shall seek to show that current analytical approaches, normative discussions and even political problems have their roots in experiences over a much longer time-span than of the post-war period alone.

This paper is exploratory in nature. It shows a somewhat uncertain mixture between normative concerns, analytical statements (based on logical and conjectural reasoning), and empirical illustrations taken eclectically from different European countries. In fact, it raises questions, rather than that it presents definite answers.

## II. Parties and historical mobilization processes

We tend to regard the development of political parties as the natural and inevitable concomitant of mass democratization processes. Yet, a closer inspection of the record reveals that the manner of democratization left a

legacy which has had an impact on discussions of the role of parties, as well as on the functioning of parties themselves, to this day.[3] We shall order our remarks in this section along the following themes: the lingering impact of anti-party sentiment; the view that parties are essentially a transient, because transitional phenomenon; assumptions that at one time "a golden age of party" existed; the impact for better or worse of developmental perspectives in the drawing up of typologies in the field of party analysis; the Rokkan-freezing proposition; and the general issue on whether parties are dependent or independent variables in relation to social changes.

### 1. Lingering anti-party sentiment

From at least two sides the legitimacy of parties has been contested in the period of mass democratization, from that of traditionalist-aristocratic doctrines on the one hand, and ultra-democratic reasoning on the other. In the first perspective there was no place for the lower orders of society to tinker with the precious institutions of authority. When parties did eventually arise, they were at most assigned a highly restricted sphere for political action – a view long buttressed by legal doctrines which resisted the very idea of a *Parteienstaat*. A denial of party from a very different perspective derived from the teachings of Rousseau and his disciples. Starting from egalitarian premises, they emphasized the need for immediate and total participation of each individual, not mediated by partial associations or delegated representatives. Hence, neither parts, nor parties or partisans, were really compatible with the expression of popular sovereignty in a *volonté générale*. The first (authoritarian) tradition and the second (ultrademocratic) tradition are curiously married in populist doctrines which emphasize the need to tie leaders and followers directly together through plebiscitary devices which are not meant to divide political society but to cement the bonds between a leader and a people postulated to be fundamentally at one.

Clearly, these varied doctrines have been unable to stem the growth of parties in all stable, competitive democracies. Yet, it is as well to remember that they often exert an influence on contemporary "critical" discussions of the role of party in modern society.

### 2. Parties as a transient phenomenon?

A number of authors have argued that parties should really be regarded as a historically specific phenomenon thrown up by a transition from a mainly "traditional" to a more "rational-legal" social order in the Weberian sense.

Parties were necessary instruments to bring new social strata into the political process, so it is argued, but once "parochials" have turned into "subjects" and better, into "citizens" in the Almond-Verba sense, parties have outlived their usefulness. Instead citizens can act directly in different social groups according to specific issues at stake, while simultaneously bureaucratic agencies and the legal system allow adequate direct action for specific goals in which party may be a hindrance rather than a help. It is as well to remember that such sentiments were already voiced by classical authors like Ostrogorski[4] and Michels[5] who turned increasingly negative towards the very idea of party, and to recognize similarities on the other hand between such views and current anti-party sentiments of neo-democratic ideologues. A useful subject for study would be to probe such similarities, and to investigate to what degree they spring from particular political situations, in particular countries and social groups.

## 3. Was there ever a "golden age of party"?

The view that parties were essentially historically-specific agencies of initial mass mobilization, allowed yet another opinion to grow up which colours many of our present-day discussions, i.e., the idea that at one time in the history of mass democratization parties could rightfully claim to be spokesmen *par excellence* for clear principles and distinct social groups which they organized and mobilized in effective social action. Yet, in the very act of their achieving success, parties could not help being corrupted, by the opportunism of office, by a blurring of social cleavage-lines which inevitably followed successful social integration of their clienteles, by internal bureaucratic sclerosis, etc. However, the idea of a "golden age" of party is likely to be myth, rather than reality – called forth by *post hoc* idealization more than by a clear inspection of actual records. At a minimum, one should again emphasize that the actual role of parties in processes of political mobilization and emancipation were very different as between parties and countries. In some countries parties which predated the processes of mass democracy adapted in forms which were often very effective for all their being far from heroic (e.g. the US party system in the time of machine politics). In yet other cases, universal suffrage was imposed from above with the clear intention to integrate voters in the State rather than in parties regarded with little sympathy (e.g., in Bismarckian Germany, and in a very different manner in a Bonapartist or Gaullist France). Again, in countries where parties did play an important role in mobilizing new groups into political life, not all parties were equally effective in different time periods, in different regions, or in relation to different social groups. At least one disadvantage of looking at parties with historicist-ideological blinkers is that it leads one to exaggerate the degree

of mass participation and mobilization, or even the salience of party and party politics, for the citizenry at large even in the time of an expanding suffrage – a factor which cannot but lead to exaggerated and prejudicial views on contemporary developments.

## 4. Developmental typologies: Duverger and Neumann

Maurice Duverger derived both his distinction between cadre parties and mass parties, and that between internally and externally created parties, from a developmental perspective – relating them to the very different positions that notables on the one hand, and lower orders in society on the other occupied during the time of mass democratization.[6] Duverger posited a contagion from the left: mass parties were the result of a successful organizational effort of left-wing parties: once such parties had penetrated fully into the representative system, the need for competition would force other parties willy-nilly to follow suit. The argument is not very much different from that of Sigmund Neumann, who saw a trend towards evolution from parties of individual representation into parties of social, or even parties of total integration.[7] Paradoxically, there has been very little empirical testing of such assumed developmental "laws" towards ever increasing mass organization. But at least since Otto Kirchheimer[8] launched his alternative forecast about the inevitable development of catch-all parties, prevailing reasoning has tended to go into the direction of a relative weakening, rather than strengthening of parties as agencies of political mobilization. Again, the Kirchheimer proposition itself has come under fire for being a postulate rather than an empirical reality.[9] Clearly, we can only solve the problem whether such unilinear trends exist or not, by comparative analysis of a longitudinal nature.

## 5. The Rokkan freezing-proposition

If one speaks of the continuing impact of past mobilization processes on present-day politics, the Rokkan freezing-proposition[10] provides undoubtedly the most famous example. Rokkan, too, analysed the establishment of parties in Western Europe in terms of successive waves of mobilization of new groups in societies in which the suffrage expanded. Once parties had successfully integrated such new groups in the political system, party systems froze along the alignments which had given rise to them initially. As political conversion of voters from one party to another is a much more difficult task than the *encadrement* of new arrivals in a political market, party systems became relatively stabilized. Rokkan's analytical approach has been confirmed by a variety of studies (e.g. Rose-Urwin;[11] see also

Pedersen on *Party lifespans*[12] who indicated that the infant mortality of new parties is much greater than the death rate of more mature parties). Yet, there are a number of caveats. In the first place, different systems differ markedly in their aggregate volatility, with some systems becoming more and others less stable over time (see data in Pedersen[13] and Maguire[14]). Secondly, the present world of competitive party systems shows a substantial number of cases which did have marked regime changes. In some of them the party system has indeed changed drastically as compared with earlier periods of history (e.g., post-1949 Germany, post-1958 France; to a different degree all political systems in Southern Europe). Thirdly, recent research suggests that once frozen-alignments are thawing increasingly in a large number of countries, albeit at a very different rate (Maguire).[15] Finally, it is one thing to suggest that the format and alignments of party systems show considerable continuity over time – this does not necessarily imply that parties within such systems are really a replica of their ancestors. Continuity may be the result as much of intelligent adaptation of parties as of a sclerosis of which they are sometimes uncritically accused.

## 6. Parties as dependent or independent variables

The point just made may be generalized. The literature knows two approaches which seem at first sight to lead to very opposite conclusions. In one view, parties are above all historical-sociological phenomena; being the product of past political and social history they are thought to be mainly a dependent variable, then as now. In another view parties are very much the independent variable. Thus approaches analysing party strategies in terms of formal models, as well as theories about rational voting, assume that both parties and voters are free to move, and in that process to determine actual outcomes. It is suggested that this apparent dichotomy runs very much into the danger of being a superficial simplification.[16] Even the most frozen of Rokkan's party systems have clearly survived not only because they were simply the products of happy history, but because their parties showed a remarkable capacity to heed – not to say: to meet – new challenges. "Stable" party systems have survived major challenges, including world wars, a long depression, a period of reconstruction and unwonted affluence, sometimes losses of colonial empires, new international alliances, processes of political integration, and a variety of domestic societal changes, which must be distinguished between changes in governmental systems, societal changes and policy problems. Clearly, then, "old" parties have indeed proved a remarkable capacity to cope, which is something rather different from saying that they are dependent variables only.

On the other hand, the formal theories of parties and voting have had to recognize many limitations in practice, including ideological positions, inertia and familiarity considerations, permissible and non-permissible coalitions, policy distances, etc. which all seem to suggest that parties are not as "independent" as earlier theories confidently held. Parties, being normatively as well as factually independent actors, have to face new political and societal problems which are themselves not *easily* controllable, yet also are not forces of unstoppable doom.

## III. Party and citizens in abstracto

We now turn to an analysis of the relation between party and citizens, each seen *in abstracto* without taking into account the simultaneous existence of more than one party on the one hand, or the presence of alternative agencies of political mobilization on the other. Our concern is with the relationship of at least the following actors: party voters, party members, party militants, and holders of representative or executive office. We shall group our (admittedly eclectic) remarks around four headings: the problem of the electoral mandate which is supposed to bind voters, party members, party activists and leaders together in one agreed party position; the role of party militants; the conflicting trends towards party bureaucratization and a new amateur politics; and the possible conflict between parties as policy-setting and mere plebiscitary agencies.

### 1. The problem of the electoral mandate

In any discussion about the linkages between elected leaders and followers, the problem of what in England is called "the electoral mandate" looms large. The specific party platform which a party offers in elections, is the basis of its electoral legitimation, and it also forms the bond between elected representatives and their parties. But it is precisely at this point that problems emerge. For to whom are elected representatives accountable, to voters who choose them, or to party members who nominate them ?

To Duverger the problem was inextricably bound up with his view about the natural development of parties: whereas cadre parties and internally created parties could easily emphasize a *trustee* concept of representation, mass parties and externally created parties inevitably came closer to *politico* and *delegate* conceptions of representation.[17] As his developmental "laws" lack genuine empirical validation, they can hardly solve our problem, however.

A contrary, equally simple explanation has been put forward by Robert

McKenzie.[18] In his view, the force of parliamentary institutions cannot but shift the centre of actual power to the parliamentary leaders of a party, whatever the rules of a party say about its internal decision-making processes. His analysis has come under fire from Labour Party critics like Miliband,[19] and seems increasingly unrealistic given contemporary developments of the British party system, where power seems to move increasingly towards extra-parliamentary actors, particularly in the case of the Labour Party.[20] Yet, explanation is complicated by the fact, that not only very different situations prevail in different parties, but also in different countries, and that we have very little really reliable information about the actual relations between elected representatives and extra-parliamentary party organs, at any one time, let alone over time.

A realistic analysis of the problem of the electoral mandate must begin from a realization, that its importance cannot be but relative. Any electoral programme, however well prepared, inevitably faces unforeseen circumstances which may cause havoc in the best intentions. Also, electoral promises find an unavoidable restraint in the realities of bureaucratic implementation and the restrictions of available resources. Moreover, whenever parties do not gain an independent majority, electoral promises will have to be adjusted to conflicting demands of coalition partners.

However, such logical relativizations do not really solve the dilemma – not least because party militants active in extra-parliamentary party organs are not likely to be passive spectators when they see an electoral platform for which they may have fought hard, being compromised. The dilemma is perhaps most easily illustrated by juxtaposing a *party government* and a *party democracy model* (see Table I).

*Table I*
*The party government and the party democracy model*

I

*The party government model*

Party leaders in office and/or representative positions should have freedom, because:

(i) they owe their constitutional mandate to the voters;

(ii) office holders and representatives have a responsability for continuous government and the national interest, beyond party;

(iii) no party can adequately foresee changing circumstances;

(iv) it is impossible and undesirable to refer all decisions back to the party; this is particularly true in case of coalition government;

(v) there must be clear, individual responsibilities for political leaders themselves, if electoral accountability has to have any real substance.

II

*The party democracy model*

Party leaders in office and/or representative positions should be accountable to extra-parliamentary party organs, because:

(i) they owe their selection, nomination and election to party;

(ii) office holders and representatives have a responsibility to enact policies, which were the basis of their programme, as established by party;

(iii) if policies must be changed, this is not only the concern of leaders, but also of the entire party;

(iv) leaders have a tendency easily to identify their desire to stay in office, even at the expense of explicit compromises on party programmes, with party interest. It is the task of party to make sure that policy choices take precedence over leaders' interests;

(v) democracy requires participatory procedures, not only periodical plebiscitary assent.

In the first model, the emphasis is on the need to preserve freedom for elected representatives to act in the parliamentary arena. In the second model, the focus is on the contrary on the need for the leaders to remain accountable to those who nominated them. Neither model is by itself "right"; jointly they illustrate the conflicting cues to which political leaders are inevitably exposed. Any realistic analysis of political mobilization should therefore take the possibility of conflict between leader appeals to voters and to party activists into account.

## 2. Leaders, militants, members and voters

Parties differ greatly, not only in the number of their members, but also in the degree of membership activity and mobilization. Yet even in parties with a large and relatively active membership, true militants form at most a relatively small minority. May's "law" of curvilinearity,[1] which concludes that activists tend towards relatively extreme ideological positions, compared to both leaders and ordinary party members and voters, has intuitive appeal and some evidence going for it. Yet much more comparative research in this area is necessary before unambiguous conclusions may be drawn. Assuming the "law" holds, how to explain the May phenomenon? Is it because militants are by definition exclusively oriented towards party, while both leaders and ordinary members or voters are subject to greater cross-pressures? Should explanations be sought in the particular milieus and occupations from which activists tend to be recruited, i.e. those who have a relatively great control over their working hours, so that they enjoy differential advantage in spending time on political action, e.g., teachers, students, employees in social organizations? Or should one rather think in terms of different positions and responsibilities, allowing for the fact that the ideologues of today are often aspirants for real power who once they attain their goal are likely to develop very different political orientations (in a process reminiscent of the Paretian theory about a natural circulation of elites, with lions turning into foxes)?

Whatever the possible explanation, we again lack reliable evidence on

whether the role of party activists within parties has of late increased, or not. It is tempting to reason in terms of a widespread effect of the 1960s, when a new generation (or rather a minority within it) was seized by ideological hopes and beliefs, and a willingness to resort to new types of political action, multiplied in its effect by (passing) media attention and (temporary) successes. It is also possible to argue in terms of new cleavages becoming salient in political perceptions, e.g.: international preoccupations veering from Cold War to North-South conflicts, the susceptibility of new affluent groups for post-materialist values; new concerns about the environment; renewed debates about the nature of capitalism; or conflicts about armaments and peace policies. It is perfectly possible to argue that on the one hand the real participation of rank-and-file members has declined in numbers and volume, and yet to insist that militants within parties have exerted an increasing pressure. But again, we have no real data, which must clearly differentiate between different parties and countries to be at all realistic.

For a proper view on the role of party militants, we should, moreover, have much more information on internal party decision-making. We have in fact remarkable little insight on a comparative basis on crucial questions such as: who appoints party leaders? How often do party congresses meet? What regulations exist on the degree of freedom for elected representatives? Who draws up, and decides on, party manifestoes? How binding are such manifestoes for candidates offering themselves for election? Who nominates parliamentary candidates? What conditions exist for renomination? Do procedures exist for recall? How binding are decisions of party congresses? Which groups in the party decide on the formation or ending of political coalitions? Obviously statutory provisions can tell us something about such matters, but if anywhere the gap between actual behaviour and institutional prescription in the field of actual party politics is large indeed.

### 3. Party bureaucratization or a new amateur politics?

The (as yet scanty) literature on party organization and participation suggests two possible trends: on the one hand a growing professionalization and bureaucratization of party organizations, and an increased role for party-controlled personnel in government employ; and on the other hand a new tendency towards selective amateur politics.

The trend towards increased professionalization could be illustrated by the following developments:

(a) elected representatives are becoming engaged full-time; the burden of parliamentary work has become such as to make a genuine combination with other employment increasingly difficult, if not impossible; a similar

tendency is now also at work at sub-national levels;

(b) members of representative bodies tend to receive increased staff support, whether from parliamentary services or from increased personal staffs;

(c) the number of party-controlled executive positions tends to increase, as more "political advisers" are appointed, "ministerial cabinets" grow in numbers, advisory agencies expand greatly, and more generally there is increasing scope for patronage appointments;

(d) in many countries there is increasing State financing of parties (whether through general grants or specific subsidies, e.g., for campaigning costs, research offices, cadre training, youth work, etc.). To the extent that parties become less dependent on membership income, this factor may cause party elites to be less interested and concerned with membership activities and sensitivities;

(e) there is an increased reliance on experts in such matters as research, relation to the media, professional campaigning, etc.

These five points are merely suggestive of possible trends. As yet we cannot substantiate two other trends suggested by Kirchheimer: the increased role of interest group representatives in parliamentary representation; and the considerably increased entry of former civil servants into politics. To the extent that parliaments are becoming assemblies of organization men, their representative role may well become more selective, if not biased. Should one argue that politics, even in its party-controlled activities, tends to provide greater scope for political careers, in that there are more positions which demand full-time *engagement*, and which are increasingly attractive in terms of future prospects, providing a relatively open and fast channel for social climbers? If so, what do such developments imply for political representation and continuity?

As against such presumed trends towards increased professionalization, there is also an alleged increase in amateur politics. Notably the great expansion of the tertiary and quarterny sectors has resulted in increased numbers of persons who have relatively free control of their own working-time. Leisure time has increased for the population generally. But whereas a great majority chooses other ways to spend it than politics, certain categories are disproportionately represented in party activities: e.g. students, journalists, university staffs and others engaged in teaching, social workers, staff personnel of interest groups, civil servants, to some degree women and retired persons. Some of these categories relate to the political process in a rather special manner (and often an ideological one). In as far as within-party activities become increasingly time-consuming, the role of those who indeed can match this development with an investment of their own time, may become even more selective. Could one, then, argue that internal party politics is increasingly being monopolized by a dialogue between the politically-employed and those whose employment or freedom

from employment permit an equal amount of political effort – with the remainder of party members being reduced to spectator roles?

A true validation of such speculations is only possible through a close analysis of actual decision-making in parties. But some information can be had from other sources: e.g. a census of positions – appointments to which are clearly controlled by party; a careful inventory of the *cursus honorum* of party elites; analysis of changes in the social background of members of parliament and other representative bodies over time, etc.

## 4. Principles, platforms and plebiscitary stratagems

Parties derive their legitimation in democratic politics from their unique role of seeking voter approval for political programmes offered in elections. In the history of many parties, programmes of principles have been major documents in political thought and development. In as far as there is a new upsurge of ideological politics, such programmes might regain high symbolic importance. The same is true for more specific election programmes. Manifestoes seem to show a tendency to increase in length over time, and to cover increasingly varied subjects. This is, on the one hand, a natural corollary of the widening of State intervention which potentially politicizes an ever-widening variety of social actions. But it also mirrors a desire on the part of party to provide a place for sectional groups in its midst, as well as outside it (whether of organized interest groups, *ad hoc* action groups, or even simple categorical "groups" thought to need attention or wooing).

On the other hand, a variety of forces presses in the opposite direction. Thus, Budge and Farlie have argued that parties tend to "compete" not on the same issues, but on only those issues which they regard as comparatively favourable to themselves.[22] Modern electioneering emphasizes a clear selectivity of issues, as well as a strong personalization of a few, if not one, party leader(s). Also, modern politics is often also characterized by a shortening of horizons: attention goes out less to long-term policies and general principles, than to winning the next election, manning particular political offices, settling pressing political issues. If this diagnosis is true, it implies that there is an inherent conflict between the perspectives of *ad hoc* electioneers and office holders, and party activists who have pressed for long statements of principles and platforms exactly as a "guarantee" for desired action. It also illustrates again the dubious value of the concept of a definite "mandate", of which party activists tend to regard themselves as the chosen guardians, but which to leaders may be no more than tactical ploys or promises.

## IV. Party systems and citizens in abstracto

So far, our review has been carried on, as if one political party, complete with *its* voters, members, militants, and leaders in different political roles, existed in isolation only. We must now shift our attention to the level of the party system, to do justice to the fact that in the actual world of politics voters and party leaders are faced with the simultaneous existence of more than one party actor. We shall divide our quick survey again into four points: the impact of the electoral system on the manner in which parties and voters interact; the extent to which citizens can use their vote to affect the political system in a decisive manner; the comparative "hold" of parties and the party system on the voters; and the possible dynamics of the party system over time.

### 1. The debate on the electoral system: constraint or not?

Since many decades two rival views have dominated the debate on the impact of electoral systems. Put succinctly: in one view, advocated above all by Hermens[23] and Duverger,[24] electoral systems create parties. In his famous indictment Hermens attributed to PR ideologization of politics, fragmentation of political will, ineffective government, and a likely plebiscitary breakthrough of extremist movements. In contrast, a single-member plurality system forced a two-party system, with moderate parties contesting the middle ground between them. Duverger further elaborated on this view, by adding a third option in run-off systems also causing a gravitation towards the centre, but retaining a multi-party system. Against such views of the decisive impact of electoral systems, Rokkan suggested[25] that as a matter of actual record it was much more a question of party systems creating electoral systems, as parties which had crystallized around the time of the advent of universal suffrage, chose electoral systems which at a minimum were likely to consolidate their position.

At least two footnotes should be placed in connection with this debate. Sartori[26] has rightly made a distinction between "strong" and "feeble" electoral systems, the first forcing a certain organization of electoral forces, and the other mainly registrating passively whatever alignments and realignments (possibly including split-offs) occur in any particular party system. In the second place, there is little doubt that electoral reforms introduced after a clear discontinuity in political regimes did have a causal effect on the later development of party systems, as was amply illustrated in post-war Germany where the 5% rule greatly assisted the growth of what Wildenmann and others have called a two-and-a-half party system. Similarly, institutional reforms including the electoral system, but also since 1962 the institution of a directly-elected president under a

run-off second ballot system, was to have a large-scale effect on the eventual restoration and simplification of the French party system in the French Fifth Republic.[27]

In fact, the debate on the role of electoral systems has been renewed in the 1960s and 1970s. On the one hand, the increased polarization in two-party Britain caused many British political scientists to challenge the Hermens-Duverger-Downs assumptions about a natural tendency of two-party systems towards centrist government and moderation.[28] On the other hand, the record of coalition government was clearly upgraded, not least by Arend Lijphart's vivid challenge[29] of the Almond concept of a "continental-European system"[30] which did not really fit the case of what Lijphart himself was to call consociational democracy systems, and also by a more general awareness through studies of European cabinets that single-party government tended to be a relatively rare phenomenon in European parliamentary systems.[31]

Such developments have caused a remarkable mirror-game among reformist writers. Those familiar with a two-party system began to reject what S.E. Finer has called "the adversary model of politics" – witness the eager rush among present-day British political scientists, otherwise of very different persuasion, to reverse the traditional statement that "England does not love coalitions". On the other hand, the bi-polar model retains its attraction for reformists in multi-party systems. In addition to drawing on Downs' indictment of the basically irrationalist features of multi-party systems,[32] the dated belief in the merit of moderate swings of the pendulum, and the new impact of a post-Gaullist France such reformers also find solace in a simplistic reading of Sartori's typologizing of polarized pluralism.

Surely, the confusion and general self-righteousness of such debates make it a matter of urgent priority for political scientists and political leaders alike to spell out their hidden norms. Or should we really accept that one should simultaneously embrace such at first sight conflicting criteria as: the wish to have both alternation in government and continuity in policy; the view that political action should bring out differences and bridge cleavages; the desire to honour electoral mandates and the recognition that government presupposes the possibility to adjust declared intentions in the light of changing circumstances; the wish to retain "liberal" elite values and also to honour the sovereign will of the mass electorate, and so on?

Turning towards a more mundane level, we now have a fair number of studies which give us rather complete information on the manner in which votes are translated into seats (e.g., Wildenmann/Kaltefleiter,[33] Douglas Rae[34]) under different institutional arrangements, including the major variable of size of district. Less satisfactory, however, is our knowledge on how different systems really affect parties and voters in the choices they

must make. Speculation and simulation are all we can go on. But there, again, we are likely to run across Rokkan's argument that parties and voters do not act in a pristine world, unaffected by historical developments, but are in fact very much determined by the manner in which party systems grew out of particular cleavages, salient at a particular historical period.

## 2. The effect of voter choice: how decisive?

European parliamentary systems do not offer the same degree of voter choice, for reasons both institutional and relating to the format of party systems.

Practically in all countries, voters choose members to the Lower House, and to various regional and local representative bodies. But even on this score countries differ in the frequency of elections, as they do also in the degree to which the voter can influence the actual composition of the Second Chamber. The importance of the Second Chamber itself in different systems shows great variation, while such a body is absent in some countries. More important is the issue whether voters have a direct say in the election of the head of State or the premier – a variable which in very different ways has proved of substantial importance in Weimar Germany, Finland, or post1 1962 France. Yet another difference lies in the extent to which systems have allowed referenda (binding or consultative). The greater the degree direct voting is allowed, the greater the potential challenge for parties. Can and should they seek to enter all contests on whatever level? Are referenda a way to short-cut parties (in a manner which may not be unwelcome to them whenever they are faced with untractable and divisive issues)? And to what degree do forms of a direct election of supreme executives strengthen a plebiscitary element in voter choice, so that parties may be becoming more dependent on leaders than leaders on parties?

At the level of the party system, the major variable is, of course, the degree to which voters are likely to bring about single-party majority governments, or at a more modest level, are able to force a substantial change in the composition of governments. We do have a somewhat sophisticated terminology to characterize the manner in which governments can change under the impact of elections, ranging from alternation, two-bloc parliamentarism, semi1 and peripheral turnover, to *Proporz*-arrangements and grand coalitions.[35] We also have full inventories of the parliamentary base and party composition of cabinets over a long time-span, as well as sophisticated theories about coalition-formation.[36] But what we do not have is clear and reliable information on a comparative basis on what different institutional and party arrangements imply for

voter perceptions, attitudes and behaviour, let alone the general legitimacy of different democratic political systems. In that sense, the generally wide gap existing between scholars interested in institutions and overall political configurations of political systems on the one hand, and specialists on voting behaviour on the other, is regrettable and should be bridged.

One study we do have, which throws some light on the relation between voters and governments, is the study of Rose and Mackie prepared within the *Recent changes* project on whether incumbency in government pays or not.[37] "Common sense" confidently suggests two opposite conclusions. On the one hand, it is argued, that each government is bound to take unpopular decisions, and eventually to alienate some of its supporters; one would therefore expect governing parties not to do well in elections. On the other hand, there is the assumption that governments work in the glare of publicity, that they can manipulate the economy once election time approaches, that they can choose a moment for elections which to them seems favourable (at least in those systems which permit such tactical dissolutions); from such propositions one would expect government parties to do well. In fact, Rose and Mackie show that the available evidence does not permit either conclusion – the electoral record of governments being too chequered. One overriding difficulty is, of course, that in many cases coalition governments fall apart at election time, with each participating party going its own way. The Rose-Mackie data suggest that coalition parties suffer rather different electoral fortunes, some winning and some losing. At a minimum, this greatly complicates a systematic analysis, because one should then be able to explain why voters for parties in one common coalition do react so differently.

## 3. The "hold" of party systems

The general debate on *The future of party government* is greatly affected by the assumption that there are clear signs of a weakening of the hold of parties on the loyalties, sympathies, or even simple interest of ordinary voters. Reliable statements on actual trends can only be made by studies of developments over time in single countries, and comparative studies between them.

Important indicators would be:

### A. *Aggregate date*

1. Data on turn-out (as assembled by Dittrich and Johansen).[38] A first inspection of overtime graphs hardly confirms a clear relation between the size of turnout and a presumed well – or malfunctioning of particular

systems (let alone, that there are rival propositions which would see in low turnout and which would see in high turnout a reason for concern).

2. Data on aggregate electoral volatility (as collected by Mogens N. Pedersen).[39] Here again, trends are far from clear, because widely differing between countries over time. At the same time, however, different studies by Pedersen, Maguire, Borre, Mayer and others do suggest an increased aggregate mobility in the 1970s in most European countries.[40]

3. Data on traditional system parties versus new parties. Such data could provide a relatively easy indicator, at least for systems which traditionally had clearly-marked system parties. But to my knowledge no conclusive data have yet come out on a comparative basis.[41]

4. Data on political fragmentation. The Rae-index has been widely used, and also severly criticized (e.g., Pedersen, Sartori).[42] Different authors have collected over-time data on a comparative basis; their studies again show substantial differences between countries, and not one linear, but a fluctuating trend for European countries as a whole (see Wolinetz and Ersson-Lane).[43]

5. Membership trends. An overtime study by Stefano Bartolini is now available for all Socialist parties.[44] Comparable analyses for other parties meet with considerable difficulty, not least because of unclear membership criteria and the greater difficulty of comparing non-Socialist parties across countries. Apparently, there is no such thing as a universal and massive decline of party memberships (although there have been drastic changes for some countries, and/or some parties).

### B. *Survey data*

Both the wide spread of electoral surveys (often modelled in one form or another on the Michigan model), and extensive comparative projects like the Verba-Nie-Kim[45] and the Eight-nation studies,[46] allow close scrutiny on a comparative basis on a variety of data relevant for the study of mobilization. One should probe available data for suitable measures, e.g. on individual volatility, party identification, political competence, political cynicism, and allegiance and legitimacy generally. But at least one major problem is that we are unlikely to find longitudinal data of sufficient sophistication to allow unambiguous diagnoses of systematic import.

## 4. The possible dynamics of party systems

There have been only few attempts to chart possible regularities in the dynamics of party systems. We have the long-standing assumption about a natural swing of the pendulum for two-party systems (and possibly also for

what Sartori calls moderate pluralist systems?). Duverger once constructed, on the basis of his interpretation of French experience, a "law" about a natural slide of parties towards the right, with new parties springing up on the left to fill the vacuum left by such movement. By far the most elaborate propositions have been worked out by Sartori, however.[47] Even then we must realize that Sartori spoke only of mechanical dispositions, rather than actual "laws". A major point at issue is the degree to which (polarized) party systems are subject to centrifugation or not.[48] More generally, there is a rarely made explicit view about the importance of a "centre" in party systems, and the extent to which the existence of one or more centre parties is likely to make for centripetal or centrifugal drives.[49] Yet, is there not a curious discrepancy between the large number of elections which provide a considerable body of data for analysis, the flourishing of a large body of abstract coalition theories, and the only very limited literature on the actual dynamics of European party systems, informed by both substantive knowledge of the party configurations of countries concerned and the high analytical skills of an author like Sartori, or Downs? More systematic work along their line, in the light of actual over-time developments of European party systems, should have a high priority in future research.

## V. Parties, citizens and other intermediate agents

Whereas in Sections III and IV we dealt with the relation between *party*, c.q. the *party system*, on the one hand and citizens on the other, as if these relations existed *in vacuo*, we must now turn to a consideration of both parties and citizens in relation to other intermediate agencies. We speak of "intermediate", because here, too, we conceive of mobilization processes as providing vertical links between the ultimate, legitimate decision-makers at the top, and the ordinary citizen at the bottom. A proper analysis will take these two different perspectives: from the top downwards (Section V), from the bottom upwards (Section VI), duly into account. We shall assume as one major hypothesis that in all countries there is a general process of increasing social differentiation, leading to a greater assortment of linkages as well as to a multiplication of sites of decision-making.

### 1. Parties and group mobilization

In an ideal-type party of social integration as described by Sigmund Neumann, parties direct themselves to specific social groups, which they seek to mobilize and incorporate in the life of the party. The characteristic

form of organization is the ancillary association, which under party auspices seeks to meet the interests of specific categories of citizens: women, youth, trade unionists, those in search of specific forms of recreation, like sports, singing or other music, etc. Parties also try to strengthen their hold on communications media, by not only providing specialized party information bulletins, but by seeking to control general newspapers and also, if at all possible, radio and television. In the same way, parties have engaged in insurance activities, in elaborate education programmes, and more generally as brokers in obtaining specific government services. As we said earlier, there are many indications that such closed worlds are breaking down, as specific party-bound organizations and media could in the long run not compete with more specialized media in an increasingly mobile and open society. But here again, one must sound a word of warning before too general a conclusion is drawn. For countries and parties again differ greatly, with forms of party "colonization" being much stronger and persistent in some countries (Italy, Austria, Ireland, Finland?) than in others.

Assuming such closed networks of party-controlled organizations are weakening, or even disappearing, what does this mean? Does it imply an over-all weakening of parties, as they become less socially "rooted" and have fewer dependable resources for mobilization? Or should one argue, that it also frees parties from historically-transient "ghetto positions", which could threaten their very existence in a period of new issues and concerns?

## 2. Parties and formal interest groups

The relation of parties to major interest groups represents one aspect of the general problem just mentioned. In a Neumann-type party of social integration, but paradoxically also in the Kirchheimer-type of "catch-all" party, stress is laid on the reciprocal advantages of network relations between parties and major interest groups. Again, we have little systematic, longitudinal information on this point – let alone that we are in a clear position to state who is "colonizing" whom, and what possible changes may have taken place in such forms of interaction. Certain over time data can be derived from systematic studies of changes in the composition of parliamentary personnel; at one time-point, the Middle-level elites project offer cross-country and cross-party information. Yet, at first sight, one might venture the hypothesis that many important interest groups are finding party a less important actor with whom and through whom to work than are ministers of whatever persuasion, as well as permanent bureaucrats.

## 3. Parties and "action groups"

Self-styled advocates of the "new politics" of action groups have adduced two patently false arguments to account for the salience of new political groups in the 1960s and 1970s.[50] One false argument is that they are a new phenomenon, as if the history of social mobilization is not replete with examples of action groups in addition to parties (which in certain cases even originated in the forms of looser voluntary societies – note as one example the emphasis on the "counter-cultural" or "peripheral" protest movements in the mobilization of the Old Left in Scandinavia). A second false argument is the presumed closed, oligarchical nature of many existing political parties, as if their history does not reveal an exceedingly low threshold for the successful entry of conscious minority groups. A number of recent studies have pointed out, that many activists in action groups are also disproportionately active within parties, often to the regret of existing power holders and believers in the merit of institutionalized politics generally. Whereas some party politicians have sought to embrace action groups warmly ("to remain in touch with the young", "to take note of the issues which clearly mobilize new groups", or even simply "to be with it"), there is little evidence that such tactics have durably strengthened parties organizationally or electorally. For parties face at a minimum two dilemmas: they must aggregate as well as articulate, and given the often extreme positions of forces demanding to be "articulated", the task of aggregation is neither easy, nor lasting, nor always electorally rewarding; and given the low cost-benefit ratios for minority groups engaged in flamboyant actions, for as long as they obtain privileged access to the political agenda, they are not likely to forfeit such benefits for the sake of party loyalties or convenience. Clearly, the relations between parties and (many) *ad hoc* action groups is therefore likely to be a thorny one, as compared towards links with more established and stable interest groups, let alone the ancillary organizations under party control discussed before.

## 4. Parties and mass media

The subject of this paragraph is one of the most complicated ones, not least because the heading covers a variety of possible relationships. On the most general level, there is the social impact of changes in the modern communications system, including not only the ever-increasing variety of printed publications, but above all radio and television. On a somewhat lower level of abstraction, there is the issue what the effects of modern media developments are on the life of parties, both in their external presentation and their internal functioning. And finally, there is the more mundane matter of the extent of actual party control over the various media and their content.

On the basis of impressionistic evidence, parties would seem to have been unable to retain the control they once may have had over parts of the daily press: post-war history is replete with financial failures of one-time strong Socialist and other *Weltanschauung*-newspapers. Control over radio and television offers a much more varied picture, with situations differing greatly between countries. But even here, the competition from alternative broadcasts seems to have seriously weakened the force of such controls as parties may have maintained or established. As compared to earlier times, few must be the groups which are still exposed mainly to a closed communications network (if ever they were).

As for the second problem, of the impact of the media on the external and internal relations of parties, again speculation must take over where concrete, detailed studies fail. One could submit the following three conjectures: First, the role of the modern media has strengthened the tendency to personalize top leaders, to give greater scope to the manipulation of issues in a selective manner, and to engage above all in short-term tactical politics. Second, the natural tendency to regard conflict rather than consensus as news is likely to be to some degree a self-fulfilling prophesy, as ardent minorities seek media attention for particularist goals. Third, the general distraction which modern mass media provide, may unfavourably affect participation by rank-and-file members, thus eroding many of the characteristics which a *parti de masse* was traditionally (and probably ideally) thought to represent. To the extent this is true, this factor strengthens the role of select groups of activists in the manner analysed earlier in this paper. (For a sophisticated attempt to relate media developments as one major variable in changes occurring in party systems to some other crucial ones, in the light of concerns about political accountability, see Sjøblom).[51]

## 5. Parties and bureaucracies

In view of the joint weight of parties and bureaucracies on actual decision-making, this relationship is probably the most crucial one. From the point of view of party leaders, the functioning of bureaucracies poses at least three challenges: (a) parties must be able to ensure effective political control and democratic accountability; (b) control over the bureaucracy can be a valuable resource, for the elaboration of specific services; (c) the bureaucracy, like other agents mentioned earlier, provides an alternative linkage mechanism between government at the top and citizens at the bottom. Seen from the point of individual citizens, the bureaucracy is one major part of "government", which to them is not always easily distinguished from parties in control. Empirical study (in the Verba-Nie tradition)[52] have emphasized direct contacting of officials as one major form of

participation in political systems. The degree of satisfaction with bureaucratic responses can therefore be a major factor in the creation and maintenance of legitimacy, from which paradoxically also individual parties, the party system and the constitutional order generally can profit.

Yet, the very term bureaucracy covers in fact an ever widening number of offices, authorities and apparatuses, which are split more and more along functional as well as regional lines. Both absolutely, and relatively in relation to all other actors, the "weight" of different bureaucracies in relation to other political actors must have increased most in the post-war world.

In the actual relation between parties and bureaucracies, one major variable, of continuing importance, consists of past State-building and nation-building processes. There is a world of difference between countries in which these processes developed largely without the prior or simultaneous establishment of powerful bureaucratic structures, and those in which such structures did play an indispensable, or at least relatively powerful, role. In countries with a consociational tradition, or one in which centralization was as much the result of judicial, aristocratic and representative processes (as in England), bureaucracies developed late and generally fell in easily with the assumptions of accountability to legitimate political authorities. A very different situation prevailed in countries like France, or Prussia, or in yet another manner Italy, where the bureaucracy remained in many ways a *corpus alienum* super-imposed on a highly localized society. To the extent, that bureaucracies live out of prel or non-democratic traditions, the problem of partisan control and political accountability is a much greater one than elsewhere.

At the same time, the processes of wide-spread diversification cannot but make havoc with earlier assumptions about one homogeneous bureaucracy, as presupposed in such different viewpoints as the Weber ideal-type, the British Civil Service, or the French *Haute Administration*. A true analysis of actual relations between parties and bureaucratic agencies must go beyond a general statement about relations existing at the top echelons of government, to lower levels both along regional and functional lines. In either case very different relations may prevail, with also rather different party actors being concerned. (Is not Italy a major illustration of this case, as the system does not depend, as so often argued, on the control over the central government by the DC-cum-minor-partners alone, but also on at least four subordinate levels of government, not to speak of the remaining *sottogoverno*, in which often other parties have a very real share ?)

One might hypothesize that *in toto* the role of bureaucratic agencies and individual bureaucrats has substantially grown as compared with those in the partisan-electoral channel. One might also point to the very substantial role that bureaucrats rather than politicians play in what are often too

easily and generically called corporate channels. One could then suggest that parties have to some extent tried to "compensate" by a variety of devices to strengthen control over and above their formal powers in government and parliament: e.g., by the multiplication of ministerial personnel; by the addition of political advisers, by a strengthening of parliamentary staffs, by the formation in some countries of ever more numerous ministerial *cabinets*, by conscious political appointments in established bureaucratic positions, by setting up new bodies outside normal hierarchical official channels, etc.

To the extent that such posited developments do indeed occur, it is a moot point whether they indeed strengthen party as an institution or not. Of old patronage has served to cement certain loyalties. But at the same time, diversification at the political level mirrors that on the bureaucratic level. Party may be a necessary channel for coveted appointments. But once arrived, political appointees need not necessarily care much about party life and tasks.

## VI. Citizens, parties and alternative agencies of mobilization

### 1. Citizens, active or not?

So far we have focused on the interaction between political actors, who are by definition active in some form or other: top elites, party militants, interest groups, action groups, the media, bureaucrats. When we now direct our attention to the citizens, activity cannot be taken for granted, as indeed the very concept of political *mobilization* suggests. In fact, there is the very fundamental issue whether citizens should always be active. On the one hand, there is the widespread assumption that citizen activity is by nature a "good". Is the history of democratic development not written in terms of the activation of ever new strata in political decision-making? Do we not, ever since Rousseau, have the whole weight of the participatory democracy literature resting on us? Is not equality of participation a must, for the citizen and the system both – because the equalities of input, the fairness of the political agenda, and of the "authoritative allocation of values" depend on citizen activity? Yet, it is as well to remember that there is also a widespread counter-literature: fears from Mill to Kornhauser[53] about a "mass society"; the proposition ever since the study on *The authoritarian personality* that lower class entry might hazard liberal elite values (e.g. Berelson,[54] Lipset[55]); the traumatic experiences of enforced political mobilization by totalitarian regimes, etc.

At a minimum therefore there is a fundamental problem of the amount or degree of political mobilization which a political system can sustain, the

nature of demands, the methods used – problems rightly posed by different models of politics, as one can note on the stress on gate-keeping and support in systems theory, on the danger of overload in communications approaches, on the need for adequate institutionalization if political system is not to fall into political decay (Huntington), etc.

This implies no less than that our subject raises issues which at the more fundamental level of analysis brings the entire armory of empirical democratic theory into play: e.g., discussions of pluralism, in developmental and contemporary perspectives, the Eckstein-theory on congruency of authority patterns,[56] the entire literature on political legitimacy. And the same holds true, on only a slightly less general level, for the literature on political socialization, participation, the voting literature, empirical studies of political legitimacy, and so forth and so on.

## 2. The potential for action

The literature on historical mobilization in Western democracies provides some data and more general speculations on the circumstances under which citizens act – or perhaps on a more general level on how people change from "parochials" and "subjects" to "citizens", in the Almond-Verba sense. Handbooks with time-series data provide a number of relevant social indicators. For our contemporary understanding the most important data are undoubtedly in the Eight-nation studies, which have made the problem of dispositions to act, and to act in different manner, the core of their theoretical and empirical concerns.

## 3. The modes of action

There is by now a large-scale debate on the modes of political participation. The Verba-Nie-Kim analysis with its emphasis on at least four independent dimensions of participation: voting, party and campaigning activities, communal action, and direct contacting – which are theoretically *and* empirically distinct – are particularly relevant. While confirming a strong bias in participation by the higher-educated, higher social status and higher income groups, they also introduce many *nuances*, depending on the type of issue or kind of political action required. Such findings must be weighed in the light of the findings of the Eight-nation study, and possibly different research results on the basis of additional survey research questions, notably on the independence of a separate protest dimension in addition to the four singled out by Verba et al. Also of great importance is the general conclusion by Verba c.s. that well-institutionalized forms of social organization (as provided by strong parties, interest

groups and bureaucratic agencies) can do much to redress the inequalities of participation as far as actual policy outputs are concerned.[57]

## 4. The channels used

Survey evidence reveals again and again that voting and direct contacting of administrative agencies are the most widespread kinds of citizen actions. On the other hand, real participation in parties is low, by whatever measure taken – even though party membership figures in relation to party voting reveal very substantial different situations, according to country or party. Measures of constancy of voting for the same party, and of party identification, can tell us something further on the degree to which parties are salient for individual citizens; to the degree that reliable longitudinal data are available we can form some impression on the degree of hold of parties on citizens.

"Groups" are more immediately relevant citizens than are parties. All citizens experience the influence of various primary groups, which the literature on political socialization and voting have singled out as being of the utmost important. Many citizens are members of associations, some of which are deliberately, some only incidentally politically active. The Verba-Nie-Kim studies have attempted to inventorize to what extent leaders and citizens perceive different groups as being actually politically relevant. Even if citizens do not, the presence of such groups, as well as their presence in them, have a substantial importance for policy-making and political attitudes.

We have discussed earlier the interaction between parties and action groups, and challenged the stereotyped view that citizens "must" turn towards the latter, because parties are insufficiently open and responsive. Verba et al. did, of course, find communal action to be a relatively independent dimension from party activities. Yet one should remember that what we conceive above all as action groups represent only a small degree of the wide range of spontaneous citizen activities studies by Verba and his collaborator. Especially, when our focus is on the most visible, most directly "political" type of action groups, we should emphasize that participation in them is highly skewed towards particular, higher-educated groups in society, who are also disproportionately active in parties. We have also data on some countries on the degree of legitimacy accorded to action groups, and to different modes of protest behaviour. Such data will need close study, if we are to have a clear view of the relative legitimacy accorded to different kinds of institutionalized actors (including party) as against less-structured ones. Just as we should not *a priori* regard parties and action groups as alternative modes of political action, so we should also not confuse all types of *ad hoc* actions with protest behaviour.

### 5. Mobilization and policies

Viewed from the perspective of the citizen some substantial problems arise on the role of parties in policy-making. Although individual citizens may mobilize on behalf of particular policies, much the larger part of policies devised or implemented are not the result of any obvious behaviour on their part. In voting, citizens may have a correct or very incorrect view on actual responsibilities for policies they might wish to reward or penalize. Citizens may or may not have particular views about how effective alternative agencies of mobilization for pressure on policy-makers may be. The current debate on whether voters vote in the light of their own economic prospects, or their view on how those in power handle economic policy, or neither,[58] should be a reminder that the feedback of policies on citizen attitudes and behaviour may be very different from what policy-makers expect. Again, there is some evidence from the Eurobarometer studies that many citizens are becoming *more* satisfied with their own situation, as the outlook for the future darkens. Clearly, levels of expectation exert a powerful independant role on the extent to which people are likely to hold politicians really accountable for policies, or their absence. What these short remarks simply re-affirm, is the urgent need to relate findings from survey research on citizen perspectives and attitudes to the policy outputs and the general functioning of political systems, over time and across countries.

## VII. Parties and other linkage actors: displacement or changing roles?

### 1. The dangers of evolutionary reasoning

In the very concept of *The future of party government* there is more than a hint that parties may be on their way out, along a foredoomed trajectory. In that respect, it is as well to remember at least the following points made in earlier sections:

(a) It is too readily assumed, without much evidence, that there was a golden age of party which is now being moaned. Should one not, at a minimum, take into account that different countries had very different degrees of party government in the past?

(b) Similarly, there is substantial evidence that parties in different countries have, if anything, been strengthened in the last quarter of a century or so. Both in France and in Germany, parties have made an impressive comeback, and parties clearly remain exceedingly important actors in many other countries.

(c) There is considerable danger in reasoning in zero-sum terms. Even

when other intermediate actors increase in importance, it does not necessarily follow that such developments go to the detriment of party, let alone that power is actually taken away from them. In as far as the overall role of government increases, one should allow for the possibility that all actors, including parties, play an increasingly important role.

(d) It may well be that the role of parties increases in some areas, while it decreases in others. One cannot really speak of "the" power of parties, without specifying power in relation to what or whom.

## 2. Some tentative generalizations

With that very substantial proviso, one might suggest that the following developments affect, at a minimum, the role of parties :

(a) there is a clear increase in the number of groups – both institutionalized and non-institutionalized ones. Most of these operate independently in the political arena, in that they attempt to influence other political actors directly – as much as they may, or may not, seek to influence party;

(b) the number of arenas for political decision-making has increased, notably in connection with what had been called "sectoral" politics;

(c) the media play probably an increased role in articulating political demands, as well as in the manner in which demands are being articulated. Parties as much as other political actors seek to adjust to this fact; Yet, the immediacy of newspaper attention gives relatively small groups new opportunities for direct approaches to power holders, without the intermediate channel of party;

(d) there is an inc[illegible] element of plebiscitary politics – both in the exposure of particular leaders, as in recourse to referenda and opinion polls. Leaders and voters are therefore, to some degree at least, more directly accessible to one another, without regular control from procedural party politics;

(e) the complexities of modern government increase the need to call on experts – a development which parties cannot balance fully by building up party research bureaus, increasing the number of parliamentary staffs at the disposal of party, or attempts to mobilize friendly experts;

(f) the increased mass of routinized decision-making, whether in bureaucratic agencies, or in relatively closed circuits of a semi-corporatist nature, inevitably escapes party control; and its aggregate effect limits the freedom of parties to set priorities, as old commitments and sectoral politics dominate over new demands;

(g) there is also a tendency for other actors to play an increasing role in political decision-making; one should note the role of the judiciary and central banks;

(h) international decision-making exerts an increasing influence, which even a more intense movement towards the building-up of party cooperation across frontiers than exists at present, cannot hope to control;

(i) given the possible weakening of party as *the* major agency in articulation and actual policy decisions, their main importance could well be in leadership selection and the long-term formulation of policy perspectives. These tasks put heavy demands on internal party mechanisms, which do not seem to be very well suited to them for the present. However, they are at least clearly within the uncontested control of parties themselves.

## Notes

* This paper, written in 1982, is very much an interim statement: it throws out conjectures and raises queries on important and complex social developments, which indoubtedly deserve and often defy systematic empirical analyses. I am grateful to Peter Mair for critical advice and editorial assistance.

[1] Giovanni Sartori, "Concept misformation in comparative politics", *American Political Science Review*, 64:4, 1970, 1033-1053.

[2] E.g., in *Citizens, elections, parties* (Oslo: Universitetsforlaget, 1970).

[3] Hans Daalder, "Parties, elites and political developments in Western Europe", in G. La Polombara and M. Weiner (eds), *Political Parties and Political Development* (Princeton, Princeton University Press, 166) 43-77, and Hans Daalder, "The comparative study of European parties and party systems: an Overview", in Hans Daalder and Peter Mair (eds), *Western European Party Systems: Continuity and Change* (London and Beverly Hills: Sage, 1983), 1-28.

[4] M. Ostrogorski, *Democracy and the Organization of Political Parties* (2 vols) (London: Macmillan, 1902).

[5] R. Michels, *Zur Soziologie des Parteiwesens in der Modernen Demokratie: Untersuchungen über die Oligarchischen Tendenzen des Gruppenlebens* (Leipzig: Klinkhardt, 1911).

[6] Maurice Duverger, *Political Parties: Their Organisation and Activity in the Modern States* (London: Menthuen, 1954).

[7] Sigmund Neumann, *Modern Political Parties* (Chicago: Chicago University Press, 1956).

[8] See Otto Kirchheimer's various articles in F. S. Burin and K. L. Shell (eds), *Politics, Law and Social Change: Selected Essays of Otto Kirchheimer* (New York: Columbia University Press, 1969).

[9] See, for example, Karl Dittrich, "Testing the catch-all thesis: Some difficulties and possibilities", in Daalder and Mair, *op. cit.* and editorial assistance, 257-266, and S. B. Wolinetz, "The transformation of Western European party systems revisited", *West European Politics* 2:1, 1979, 4-28.

[10] Rokkan, *op. cit.*, chapter 3.

[11] Richard Rose and Derek Urwin, "Persistence and change in Western party systems since 1945", *Political Studies* 18:3, 1970, 287-319, and Derek Urwin and Richard Rose, "Persistence and disruption in Western party systems between the wars", paper presented to the World Congress of the International Sociological Association, Varna, 1970.

[12] Mogens Pedersen, "Towards a new typology of party lifespans and minor parties", *Scandinavian Political Studies* 5:1, 1982, 1-16.

[13] Mogens Pedersen, "Changing patterns of electoral volatility in European party systems, 1948-1977: Explorations in explanation", in Hans Daalder and Peter Mair, *op. cit.*, 29-66.

[14] Maria Maguire, "Is there still persistence? Electoral change in Western Europe, 1948-1979", in Hans Daalder and Peter Mair, *op. cit.*, 67-94.
[15] *Ibid.*
[16] For a more extensive discussion, see Hans Daalder, "The comparative study of European Parties and party systems", *op. cit.*
[17] These terms are taken, of course, from J. C. Wahlke et al, *The Legislative System: Explorations in Legislative Behaviour* (New York: Wiley, 1962).
[18] *British Political Parties* (London: Heinemann, 1955).
[19] Ralph Miliband, "Party democracy and parliamentary government", *Political Studies* 6:2, 1958, 170-174; see also Robert McKenzie, "Policy decisions in oppositions: a rejoinder". *Political Studies* 5:2, 1957, 176-182.
[20] See S. E. Finer, *The Changing British Party System, 1945-1979* (Washington DC: American Enterprise Institute, 1980).
[21] John D. May, "Opinion structure of political parties: the special law of curvilinear disparity", *Political Studies* 21:2, 1973, 135-151.
[22] Ian Budge and Dennis Farlie, "Party competition - selective emphasis or direct confrontation: An alternative view with data", in Hans Daalder and Peter Mair, *op. cit.*, 369-404.
[23] F. A. Hermens, *Democracy or Anarchy? A Study of Proportional Representation* (South Bend: University of Notre Dame Press, 1941).
[24] Maurice Duverger, *L'Influence des Systèmes Electoraux sur la Vie Politique* (Paris: Colin, 1950).
[25] Stein Rokkan, *Citizens, Elections, Parties, op. cit.*, chapters 3 and 4.
[26] Giovanni Sartori, "European political parties: the case of polarised pluralism", in La Palombara and Weiner, *op. cit.*, 137-176, and Giovanni Sartori, "Political development and political engineering", *Public Policy*, vol. 17, 1969, 261-298.
[27] For an excellent discussion of institutional factors such as the impact of electoral systems, direct election of the chief executive and federalism, see Robert A. Dahl, "Some Explanations", in Robert A. Dahl (ed), *Political Oppositions in Western Democracies* (New Haven: Yale University Press, 1966), 348-386.
[28] Hans Daalder, "In search of the centre of European party systems", *The American Political Science Review*, 78:1, 92-109.
[29] Arend Lijphart, "Typologies of democratic systems", *Comparative Political Studies* 1:1, 1968, 3-44.
[30] G. A. Almond, "Comparative political systems", *Journal of Politics* 18:3, 1956, 391-409.
[31] For an early example, see Jean Blondel, "Party systems and patterns of government in western democracies", *Canadian Journal of Political Science* 1:2, 1968, 180-203.
[32] Anthony Downs, *An Economic Theory of Democracy* (New York: Harper and Row, 1957).
[33] Rudolf Wildenmann, Werner Kaltefleiter and U. Schleth, "Auswirkungen von Wahlsystemen auf das Parteien und Regierungssystem der Bundesrepublik", in Erwin Scheuch and Rudolf Wildenmann, *Zur Soziologie der Wahl* (Köln: Opladen, 1965).
[34] Douglas Rae, *The Political Consequences of Electoral Laws* (New Haven: Yale University Press, 1971).
[35] Hans Daalder, "Cabinets and party systems in ten smaller European democracies", *Acta Politica* 6:3, 1971, 282-303.
[36] The best general survey is still that of Abram de Swaan, *Coalition Theories and Cabinet Formations* (Amsterdam: Elsevier, 1973).
[37] Richard Rose and T. T. Mackie, "Incumbency in government: asset or liability?", in Hans Daalder and Peter Mair, *op. cit.*, 115-138.
[38] Karl Dittrich and Lars Nørby Johansen, "Voting turnout in Europe, 1945-1978: myths and realities", in Hans Daalder and Peter Mair, *op. cit.*, 95-114.
[39] Mogens Pedersen, "Changing patterns of electoral volatility", *op. cit.*

[40] Mogens Pedersen, "Changing patterns of electoral volatily", *op. cit.*; Maria Maguire, *op. cit.*; Ole Borre, "Electoral instability in four Nordic countries, 1950-1977", *Scandinavian Political Studies* 13:2, 1980, 141-171; Lawrence C. Mayer, "A note on the aggregation of party systems", in Peter H. Merkl (ed), *Western European Party Systems: Trends and Prospects* (New York: the Free Press, 1980), 515-520.

[41] Though it should be noted that work in this area has been carried out in the block-volatility analysis by Stefano Bartolini and Peter Mair. For a preliminary outline of their approach, see their "Report on a new framework for the analysis of changes in Western European party systems", paper presented to the Workshop on the Future of Party Government, ECPR Joint Sessions, University of Aarhus, 1982, see this volume, chapter...

[42] Mogens Pedersen, "On measuring party system change: a methodological critique and a suggestion", *Comparative Political Studies* 12:4, 1980, 387-403; Giovanni Sartori, *Parties and Party Systems: a Framework for Analysis, Vol. 1* (New York: Cambridge University Press, 1976), 300-319.

[43] S. B. Wolinetz, *op. cit.*; Svante Ersson and Jan-Erik Lane, "Democratic party systems in Europe: dimensions, changes and stability", *Scandinavian Political Studies* 5:1, 1982, 67-96.

[44] Stefano Bartolini, "The membership of mass parties: the Social-Democratic experience, 1889", in Hans Daalder and Peter Mair, *op. cit.*, 177-220.

[45] S. Verba, N. H. Nie and J. Kim, *Political Participation and Political Equality: A Seven-Nation Comparison* (New York: Cambridge University Press, 1979).

[46] S. H. Barnes, Max Kaase, et al, *Political Action: Mass Participation in Five Western Democracies* (London and Beverly Hills: Sage, 1979).

[47] Giovanni Sartori, *Parties and Party Systems*, op. cit.

[48] See Paolo Farneti, *The Italian Party System, 1945-1980*, (London: Frances Pinter, 1985).

[49] Hans Daalder, "In search of the center", *op. cit.*

[50] Hans Daalder, "Parlement tussen Politieke Partijen en Actiegroepen", *Acta Politica* 16:4, 1981, 481-490.

[51] Gunnar Sjøblom, "Political change and political accountability: a propositional inventory of causes and effects", in Hans Daalder and Peter Mair, *op. cit.*, 369-404.

[52] Verba, Nie and Kim, *op. cit.*; S. Verba and N. H. Nie, *Participation in America: Political Democracy and Social Inequality* (New York: Harper and Row, 1972).

[53] William Kornhauser, *The Politics of Mass Society* (London: Routledge and Kegal Paul, 1960).

[54] B. Berelson, P. Lazarsfeld and W. McPhee, *Voting* (Chicago: Chicago University Press, 1954).

[55] S. M. Lipset, *Political Man* (New York: Doubleday, 1960).

[56] H. Eckstein, *Division and Cohesion in Democracy: A Study of Norway* (Princeton University Press, 1966), ch. 2.

[57] Verba, Nie and Kim, *op. cit.*

[58] See, for example, D. R. Kinder and R. D. Kiewiet, "Economic discontent and political behaviour", *American Journal of Political Science* 23:3, 1979, 495-527 and, by the same authors, "Sociotropic politics: the American case", *British Journal of Political Science* 11:2, 1981, 129-161.

# La recherche en sciences politiques et sociales comme corps de metier: quelques observations comparatives sur le plan international

BERND MARIN
Vienne/Florence

Les réflexions ci-après se rapportent exceptionnellement à nous-mêmes. Elles portent sur les conditions et particularités de notre travail professionnel en tant que spécialistes des sciences sociales. L'idée de base est la plus simple que l'on puisse imaginer: espérer des révélations de l'application d'une perspective qui, tout en nous étant familière sur le plan professionnel, n'a guère été appliquée à notre propre métier et qui est à cet égard inhabituelle et non coincidente. Dans cette perspective la production de savoir dans la recherche en sciences politiques et sociales est analysée, comme tout autre travail, également sous des aspects de production et de marché. Au risque de susciter des interprétations trompeuses, on pourrait qualifier cette façon de voir de "matérialiste". Certes, elle ne permet pas d'expliquer toutes les caractéristiques de notre corporation, mais *sans* elle on ne pourra pas comprendre grand-chose à notre métier, y compris certaines curiosités devenues si familières que nous ne les percevons guère plus comme étranges.

Tout aussi rares sont par conséquent les essais d'analyse de cette sorte. Alors que nous concevons par exemple sans hésitation les médecins dans l'exercice de leur activité comme des personnes pratiquant le métier de la santé, des contraintes de distinction de notre profession s'opposent aux prétentions qui consistent à nous percevoir avec tout aussi peu d'illusions. Déjà les catégories comme "services" ou "métier" font à la rigueur penser aux épiciers ou aux ingénieurs civils, aux conseillers en placement ou aux couturiers, aux agents d'assurances ou aux prédicateurs du salut, aux courtiers, aux masseurs ou à quelque section de la branche psychologique telle qu'elle est représentée par des thérapeutes de groupes ou des psychanalystes. Or, dans les sciences sociales et dans la recherche être classé sous "les services", au sens employé par Renner ou Dahrendorf, équivaut justement à l'exclusion symbolique du monde académique: citons come exemples l'étude (commerciale) des marchés ou la statistique sociale gou-

vernementale. Cela, tout en étant inexact, n'est nullement absurde, mais représente une importante stratégie de dominance et de distinction *à l'intérieur* du champ social de la recherche en sciences sociales.

Les observations qui suivent visent un certain nombre de caractéristiques structurelles et d'éléments dynamiques de ce domaine-là qui a acquis une importance accrue en tant que recherche en sciences sociales orientée vers la politique (la pratique, les problèmes, l'application), dénommée ci-après RSSP. Pour éviter des malentendus terminologiques il y a lieu de souligner d'avance qu'une RSSP axée sur les problèmes ne doit ni être réduite à cette variante à la mode qu'est la *policy research* (policy analysis, policy studies etc.) ni être mise sur le même plan que la procuration d'informations commerciale ou liée au mandant. A la différence de la *policy research* américaine qui porte sur l'élaboration, la mise en oeuvre et l'évaluation de programmes politiques publics spécifiques – et qui, par conséquent, occupe par exemple dans le domaine du *technology assessment* avant tout les spécialistes en sciences physiques et naturelles, les techniciens ou les médecins – la recherche en sciences sociales orientée vers la politique est de par son concept plus étendue tout en étant limitée par la discipline. Ainsi la RSSP présente non seulement de nombreuses nuances d'importance sur le plan politique et pratique, mais est par tradition également productrice non commerciale de savoir ayant souvent un fond administratif, de réforme sociale ou de critique de la société.

En Allemagne par exemple cela va des vieilles *Polizeiwissenschaften* a) du XVIIè et du XVIIIè siècles auxactuelles enquêtes menées par le *Wissenschaftszentrum Berlin* ou les Instituts de sciences économiques et sociales du *Deutscher Gewerkschaftsbund* (Confédération syndicale allemande), en passant par les travaux du *Verein für Sozialpolitik* b) au XIXè siècle ou ceux de l'*Institut für Sozialforschung* c) de Francfort réalisés au cours de la période précédant le nazisme. En Autriche les premières RSSP se retrouvent dans les études de *Kathedersozialisten* d) tel Eugen von Philippovich, dans les enquêtes de l'*Arbeitsstatistisches Amt* e) auprès du Ministère du Commerce à Vienne, dans celles menées par les inspecteurs du travail et de la main d'oeuvre et les inspecteurs d'usines, dans les travaux de la *Kammer für Arbeiter und Angestellte* f) à Vienne réalisés par Käthe Leichter ou dans ceux menés à bien par la *Wirtschaftspsychologische Forschungsstelle* g) de Lazarsfeld. Or, au sens large de la RSSP ne correspondent pas seulement les recherches axées directement sur l'application, comme celles menées par l'*Österreichisches Institut für Konjunkturforschung* (par la suite:

a) sciences policières
b) Association de politique sociale
c) Institut de recherche sociale
d) partisans de la doctrine économique allemande de la fin du XIXè

siècle préconisant l'intervention de l'Etat dans la vie sociale en vue d'atténuer les différences de classes
e) Office des statistiques du travail
f) Chambre professionnelle des ouvriers et employés
g) Centre de recherche de psychologie économique

*Institut für Wirtschaftsforschung*) a), fondé par Ludwig von Mises (et dirigé entre autres par Friedrich von Hayek et Oskar Morgenstern), ou, dans un autre domaine politique, les nouveaux modèles de la *Fürsorgeerziehung* b), créés par August Aichhorn, ou la *Schulgemeinde* c) de Siegfried Bernfeld; l'étude de Paul F. Lazarsfeld intitulée *Die Arbeitslosen von Marienthal* d) est tout autant une RSSP que son oeuvre *Jugend und Beruf* e) qui a trouvé une application directe dans la réforme de l'enseignement introduite par Glöckel. Même de grandes fondations privées ont contribué à limiter la dépendance commerciale et politique de la RSSP: citons comme exemples aux Etats-Unis les fondations *Russell Sage, Carnegie, Rockefeller* et (après la Deuxième guerre mondiale) *Ford.*

Dans une première tentative nous essayerons donc d'élucider la question de savoir pourquoi une recherche en sciences sociales orientée vers la politique, à moins d'être réduite à des *policy studies*, ne pourra être analysée exhaustivement sous les seuls aspects de marché.

a) Institut autrichien de conjoncture
b) éducation surveillée au titre de l'aide sociale à l'enfance et à la jeunesse
c) Commmunauté scolaire
d) Les chômeurs de Marienthal
e) La jeunesse et le métier

## Champ et marché de la recherche en sciences sociales

Même si une partie – qui ne cesse de croître depuis des décennies et qui désormais détient quasiment partout la première place – des fonds publics destinés à la recherche sur commande va aux instituts de caractère commercial (par exemple *Prognos* en Suisse, *Battelle*, *Emnid*, *Allensbach*, *Infratest* ou *infas* en République fédérale d'Allemagne, IFOP et SOFRES en France, *Fessel* et IFES en Autriche), l'information produite par ces derniers ne correspond qu'à un pourcentage mineur de la production totale de savoir en matière de RSSP. Premièrement de plus en plus d'instituts d'intérêt public à but non lucratif se font concurrence sur le marché de la recherche financée par l'Etat – cela va des grands instituts de recherche économique et conjoncturelle en RFA et en Autriche jusqu'à une multitude de petits centres de recherche extra-universitaires.

Deuxièmement et surtout, une grande partie de la RSSP ne s'alimente

pas du tout ou du moins non pas exclusivement ou non pas de façon prépondérante sur le marché: champ et marché de la RSSP ne coincident nullement.

La production de savoir non axée sur le marché commence naturellement avec la "production propre" de dissertations ou de thèses à l'université et va jusqu'à "l'autarcie" académique des enseignants titulaires. Or, dans l'ère des "universités de masses" la capacité universitaire de recherche est en général jugée comme accusant une forte régression et comme n'étant souvent plus concurrentielle. A côté des centres de recherche extra-universitaires ou para-universitaires existe donc un nombre croissant d'organismes de recherche qui dépendent du gouvernement ou des administrations, l'exemple le plus marquant en RFA étant l'*Institut für Arbeitsmarkt- und Berufsforschung* (institut de recherches portant sur le marché de l'emploi et les débouchés) qui relève du *Bundesamt für Arbeit* (l'Office fédéral du Travail) à Nuremberg – un modèle qui progresse dans toujours plus de domaines et de niveaux politiques (Länder, communes; branches). De même ces prestations propres – plus ou moins professionnelles – de la RSSP ne restent pas limitées à l'administration publique: les banques d'émission, les groupements d'intérêts, les partis politiques, les grandes entreprises publiques et privées ainsi que, depuis peu, les comités d'action et les mouvements sociaux sont passés à des mesures "d'autoproduction" et "d'autarcie" également dans le domaine de l'expertise en matière de sciences (sociales). J'ignore s'il existe des estimations relatives à ce secteur – extérieur au marché – de la production de savoir en matière de RSSP, mais la part détenue par celui-ci dépasse sans doute de loin le quart attribué aux instituts non commerciaux et aux universités sur l'ensemble des fonds publics employés pour la recherche effectuée sous contrat sur le marché de recherche dans le domaine de la RSSP (Lutz 1977, p. 21).

Les tendances générales que je viens d'esquisser cachent d'une part une évolution qui sépare toujours plus nettement le champ et le marché de la RSSP, et d'autre part des interpénétrations nouvelles dans le sens par exemple d'une commercialisation accrue des recherches universitaires initialement extérieures au marché. A cet égard les déplacements de fonctions au détriment des universités surchargées sont d'une importance capitale: alors que les tâches d'enseignement passent aux instituts post-universitaires, aux "écoles d'élite" telles l'*Ecole Polytechnique*, l'*Ecole Normale Supérieure*, les Ecoles d'administration et de gestion ou les *Business-Schools*, les capacités de recherche se déplacent rapidement vers des instituts commerciaux, des instituts dépendant de l'Etat et surtout des instituts indépendants extra-universitaires. Même des universités de premier ordre ont dû céder des tâches de recherche à des instituts extra-universitaires: à Paris par exemple *Nanterre*, *Vincennes* et la *Sorbonne* au *Centre National de la Recherche Scientifique* (CNRS) ou à l'*Ecole de Hautes Etudes en Sciences Sociales* (EHESS), notamment aux Centres de recherche à la *Maison des*

*Sciences de l'Homme*; la *Georgetown University* de Washington au *Center for Policy Studies* ou à l'*American Enterprise Institute for Public Policy Research* (AEI), ou bien la *Stanford University* à la *Hoover Institution*; toutes les universités de Berlin au *Wissenschaftszentrum Berlin* (WZB) et les universités viennoises à l'*Institut für Höhere Studien* (IHS) et à d'autres centres de recherche extra-universitaires. Que signifie cette perte générale de fonctions essuyée par les universités dans le domaine de la recherche? Comment s'explique la suprématie de plus en plus écrasante des instituts extra-universitaires, para-universitaires et postuniversitaires dans la RSSP?

Dans le secteur RSSP du marché, où une demande stable et permanente existe avant tout sur le plan des analyses des marchés et des sondages de l'opinion publique et en ce qui concerne les produits de série technicisés et les pronostics économétriques, la forme de production supérieure des instituts extra-universitaires est évidente. Un personnel de recherche plus nombreux et un personnel technico-administratif bien plus nombreux, une spécialisation thématique plus poussée, la division du travail, l'introduction de procédés de routine, la formalisation et la collectivisation de la production de savoir, l'abandon de concepts et de méthodologies complexes en faveur de technologies simples, prêtes à l'emploi en peu de temps, accompagnées d'un meilleur équipement matériel – voilà les éléments qui permettent de produire, dans des délais imposés, des données macroscopiques de masse – une chose tout à fait impossible pour les instituts universitaires. Or, la recherche universitaire vise, déjà de par sa structure, un autre savoir: Procurés par l'intermédiaire de critères académiques de carrière prescrivant la concurrence de qualité par une différenciation des produits selon "l'originalité" etc., ce sont par principe les projets individuels innovateurs qui l'emportent sur la production standardisée de masse dans le domaine de la procuration d'information.

C'est donc avec d'autant plus d'acuité que se pose la question de savoir pourquoi la recherche universitaire a souvent perdu son avance même dans le domaine de la RSSP théorique.

Des "capacités" et des "dons" inférieurs n'en peuvent être la cause. Si l'on part du fait maintes et maintes fois confirmé que la renommée des instituts est identique à celle de leurs directeurs (Crawford et alii, 1976, pp. 20-22), de nombreux départements universitaires présentent un profil excellent.

Or, si l'on regarde de plus près, on découvre que ceux qui occupent les positions de tout premier rang ont souvent fourni leurs exploits en dehors des universités: au cours d'années sabbatiques, en tant que chercheurs visiteurs ou membres d'institutions de recherche para-universitaires etc. Comment cela s'explique-t-il? La référence – facilement concevable – à la libération des lourdes tâches d'enseignement imposées par les actuelles universités n'est tout au plus que la moitié de la vérité. Cela s'appliquerait avant tout aux centres de recherche commerciaux ou gouvernementaux

dont la production sur le plan de la RSSP est certes impressionnante du point de vue de la quantité mais reste dans le meilleur des cas un "travail de déblaiement" au sens employé par Kuhn (1970) sans pouvoir indiquer une quelconque direction au développement scientifique. Des travaux ouvrant des perspectives, s'il en est, viennent surtout des ateliers d'instituts sachant combiner les avantages des universités traditionnelles à ceux des centres de recherche modernes empiriques.

Des établissements de recherche tels le CNRS ou l'EHESS, le *Wissenschaftszentrum Berlin* ou l'*Institut Universitaire Européen (IUE) de Florence*, la RAND *corporation*, la *Brookings* ou l'AEI sont la plupart du temps mieux équipés sur le plan financier et personnel que les instituts d'université, ce qui leur permet d'entamer des études empiriques même de majeure envergure sans soutien extérieur et sans subir l'influence de tiers. Au lieu d'exploiter la capacité de travail d'étudiants, pour qui c'est une question existentielle et qui souvent ne sont pas payés (bien qu'ils préparent de plus en plus souvent les publications de leurs professeurs et directeurs de thèse), on engage des "chercheurs temporaires" rémunérés. Dans la mesure où des tâches de formation postuniversitaires sont assumées, elles servent le *seasoning* et le *learning by doing* des candidats au doctorat et des *post-docs* tout autant que le défi critique des chercheurs eux-mêmes. Aussi n'est-il bien surprenant que des établissements qui, à l'origine, furent de purs instituts de recherche, tels l'INSEE ou la RAND *corporation*, assument très volontiers également des fonctions de formation supérieure (l'ENSAE ou le RAND *graduate institute*). Le vieux modèle respectable d'une combinaison fructueuse de l'enseignement et de la recherche persiste de nos jours surtout *en dehors* des universités et en dehors des manufactures de recherche qui dépendent des clients.

Aussi d'autres caractéristiques des universités traditionnelles ne se retrouvent-elles quasi plus que dans les instituts de recherche para-universitaires: libération de la pression d'être tenu par des délais, travail et inspirations interdisciplinaires, moyens matériels suffisants ainsi que sécurité de l'emploi et modèles de carrière – du moins pour ceux qui, en tant que chercheurs bénéficiant d'un emploi fixe, vivent du beau côté du monde du travail. Alors que de nouvelles stratifications de classe naissent ici du mode de fonctionnement du marché du travail des spécialistes en sciences sociales[1], des problèmes analogues du marché du travail au sein des universités ont abouti, en raison de l'*age lump phenomenon* si souvent évoqué, à des développements erronés cumulatifs des structures corporatives et à un recul correspondant de la productivité. Cela vaut, bien entendu, uniquement pour ceux des pays européens où – contrairement au *promotional system* dans les universités américaines – la dynamique de la concurrence capitaliste et de l'évaluation du rendement individuel a été bloqué par des pyramides d'emplois figées sur le plan des corporations.

Presque partout en Europe occidentale l'augmentation explosive du

nombre d'étudiants a entraîné un accroissement correspondant mais deséquilibré du nombre de postes permanents et, par conséquent, une restructuration radicale du corps enseignant et du personnel de recherche: le nombre de postes de professeurs et surtout le nombre de chaires est, en dépit de l'expansion générale, resté inchangé[2]. L'absence d'une offre élastique de chaires résultant des titularisations et d'une faible mobilité a, conjointement à l'expansion rapide des positions réservées aux jeunes nouvellement entrés dans la vie professionnelle, entraîné une distorsion de la pyramide des âges, l'écroulement des schémas traditionnels d'autorité et des carrières et, de ce fait, des tensions et des décalages au sein des facultés. La compétition capitaliste, basée sur le rendement et visant l'éviction mutuelle, dans les instituts extra-universitaires (ou dans les universités américaines) est remplacée de plus en plus souvent, dans les universités européennes, par le principe d'une usucapion corporative des postes sur la base du *first come first served*, mélange du principe du bénéfice de l'âge et d'une prime de bonne conduite, comportant l'établissement de hiérarchies jusqu'ici inconnues entre des chercheurs qui ont plus ou moins le même âge et les mêmes qualifications. "L'ouverture" et la "démocratisation" propagées des universités a provoqué (contre tout dessein et discernement) *à l'intérieur* des universités une accentuation radicale – et un dysfonctionnement – d'inégalités et de hierarchies, et *à l'extérieur* des universités un retranchement paralysant envers toute concurrence externe.

C'est ainsi que dominent, probablement plus que jamais, des critères ignorant le rendement dans la sélection sociale et les compétitions: revendications professionnelles corporatistes de caractère militant (visant une pragmatisation générale au détriment de la nouvelle génération et se présentant parfois, malgré leur objectif réactionnaire, sous un manteau de syndicalisme "progressif"); retranchement vis-à-vis de l'extérieur (lors de l'attribution de cours et de chaires et pour les recrutements); mobilité régressive; renforcement du caractère tribal de la formation de groupes (de familles, de clans, de cliques, d'"écoles") scientifiques et consolidation des milieux de la sous-culture, souvent accompagnés d'une orientation politique; tentatives de conversion de ressources non scientifiques dans les sciences etc. Une réaction de plus en plus fréquente à ces tendances (forcément tabouisées) à la décadence (tendances qui, au niveau européen, se manifestent peut-être le plus fortement en France) est la défection: les universités ne sont plus que des lieux de passage et non l'objectif de carrière des hommes de science qui visent le *Collège de France* ou un institut *Max Planck* ou le *Centre for the Study of Public Policy* à Glasgow ou bien d'autres institutions comparables.

Les universités accordent un "capital de départ" sous forme de licences et de titres et constituent parfois également une "assurance" contre les risques de l'emploi à l'expiration des contrats avec des centres de recherche extra-universitaires, ou lors d'une "contraction assainissante" de

tels établissements de recherche quand leur conjoncture politique fléchit. Lorsque celles-ci ne sont pas, comme en France, dotées de postes de titulaires, les instituts extra-universitaires exigent pratiquement, comme contre-assurance, l'appartenance – temporairement suspendue – à une université, en particulier en Italie: ce n'est qu'ainsi que le capital humain accumulé représenté par les chercheurs peut survivre à une non-survie de leurs instituts de recherche; il suffit de penser à l'*ISAP (Istituto per la scienza dell'amministrazione pubblica)* ou au cycle d'instituts formé du CNDPS *(Centro nazionale di prevenzione e difesa sociale)*, de l'ILSES *(Istituto Lombardo per gli studi economici e sociali)* et de l'IRER *(Istituto regionale di ricerche della Lombardia)* à Milan qui se partageaient un certain nombre de chercheurs renommés.

Le même schéma vaut encore plus pour les collaborateurs des comités consultatifs politiques où seules des appartenances fixes à des institutions scientifiques permettent un va-et-vient continu – et auparavant la légitimation des nominations mêmes: qu'il s'agisse du *Council of Economic Advisers* aux Etats-Unis, de la *Paritätische Kommission* en Autriche, du *Deutscher Bildungsrat*, du groupe chargé du Projektgruppe Regierungs- und Verwaltungsreform (projet de réforme gouvernmentale et administrative), de la Kommission für wirtschaftlichen und sozialen Wandel (Commission pour le changement économique et social) en RFA, une acitivité, qu'elle soit honorifique ou rémunérée, n'y pourra être exercée que sur la base de "points d'appui" académiques des scientifiques. Cela vaut d'autant plus pour les "plaques tournantes" entre sciences, administration et politique, telles la *Brookings*. Là où la politique ne peut assurer l'environnement autonome de la recherche scientifique, elle doit alimenter directement une RSSP visant l'application: les meilleurs exemples dans ce contexte sont les différents fournisseurs du Congrès des Etats-Unis tels le *Congressional Research Service* (CRS), l'*Office of Technology Assessment* (OTA) ou le *General Accounting Office* (GAO) (Mosher, 1979) avec des milliers de collaborateurs. Or, une telle auto-fourniture en matière de RSSP de la part d'institutions non scientifiques peut très bien aboutir à des résultats respectables du point de vue académique, voire à des exploits scientifiques de premier ordre: l'unité de recherche économique de la *Banca d'Italia* passe depuis longtemps pour la meilleure équipe dans le domaine des sciences économiques orientées sur la politique en Italie[3].

Les développements diamétralement opposés des universités, des "écoles d'élites", des instituts commerciaux dépendant de l'Etat, des commissions et des comités consultatifs en matière de politique en sciences sociales et des instituts de recherche extra-universitaires, para-universitaires et post-universitaires s'accompagnent d'une multitude de hiérarchisations sociales à l'intérieur du champ de la RSSP. A la vieille hiérarchisation *entre* les disciplines accordant presque partout plus de prestige (et d'influence) à l'économie qu'à la sociologie et plus de prestige à la sociolo-

gie qu'aux sciences politiques, s'ajoutent de nouveaux différentiels *à l'intérieur* des disciplines: Des domaines de recherche d'une signification initiale "mineure" tels la sociologie de l'organisation, du travail ou de l'industrie prennent le dessus sur les vieilles matières respectables que sont la sociologie de la religion, de l'éducation et du savoir (Pollak 1978, p. 26). De même les universités et les institutions de recherche sont hiérarchisées selon leur cote en vue d'ultérieures transactions sur le marché de l'emploi: comme à une bourse invisible les cotes montent et descendent et déterminent quels collèges sont à un certain moment *hot* (en vogue), variant bien entendu selon qu'il s'agit de *law schools*, de *business schools* etc.[4].

Or, les fluctuations et les critères de cote sont eux-mêmes également différenciés (et hiérarchisés) en fonction de l'appartenance à une couche et des ambitions de carrière: un futur "grand cadre" de l'*ENA (Ecole Nationale d'Administration)* trouvera peu attrayante une carrière aboutissant à une chaire d'enseignement théorique à *Normale Sup*, poste de rêve en quelque sorte pour les intellectuels français. Enfin les hiérarchies de cote, homogènes du moins sur le plan sous-culturel, indiquent *entre les institutions* une espèce de *mobilité-fuite* d'une situation de domination insupportable au sein de chacune de ces institutions: les chercheurs qui travaillent en *free-lance* ou sous contrat temporaire visent un poste protégé d'assistant universitaire, alors que ceux parmi les scientifiques relativement jeunes qui n'arrivent pas à se faire à l'idée d'une soumission (ou d'une ascèse dans la recherche) à vie et possèdent des qualifications suffisantes, fuient les universités pour entrer dans les instituts de recherche indépendants. Vis-à-vis de cette option *exit* de la mobilité-fuite il n'existe, dans la recherche rationalisée à grande échelle, qu'un nombre décroissant de possibilités offensives prometteuses: la position académique de tête des instituts de recherche indépendants reflète elle aussi ce changement paradigmatique en sciences sociales qui définit de formes dépassées (ou très limitées) de la production de savoir en matière de RSSP toute théorie dépourvue d'expérience et toute expérience dépourvue de théorie – rehaussant ainsi les performances qui représentent une combinaison de coopération collective et de synthèse individuelle. Les développements esquissés démontrent comment le champ social de la recherche en sciences sociales orienté vers les problémes est de plus en plus marqué par des aspects de marché (et de domination politique ou de bureaucratisation), sans pour autant pouvoir être réduit à des institutions ou à des marchés.

## La recherche en sciences politiques et sociales – un métier caractérisé par la mise à disposition de capacités?

La production de la recherche en sciences sociales, il est vrai, ne s'explique pas suffisamment par les aspects d'utilisation de rentabilisation et de

marché. Dans la mesure où elle comporte cependant des éléments de marché elle peut être considérée comme une sorte de *service* ou *métier de permanence* ou bien comme une espèce "d'industrie caractérisée par la mise à disposition de capacités".

Abstraction faite des enquêtes de routine et des pronostics économétriques, c'est la "fabrication unitaire" qui y est typique, s'agissant de prestations plus ou moins "uniques" à la demande/sur l'ordre/avec le soutien d'un client collectif. Il faudra encore revenir sur la question de savoir dans quelle mesure les clients de la RSSP (à la différence, par exemple, de la clientèle des dentistes ou des conseillers financiers) sont enfin capables de définir leur "problème". La RSSP, tout en connaissant des dépassements de délais (non punis!), ignore des termes de livraison assez longs tout autant qu'une production destinée à "être stockée". Cette forme de production, consistant en la mise à disposition, quasiment sur appel, de capacités, comporte de grands risques relatifs aux capacités sur ce plan: elle nécessite un degré élevé de flexibilité dans l'emploi de la main-d'oeuvre et du capital (comme c'est par exemple le cas dans le bâtiment, la construction navale ou lors de prestations – personnelles – de service), alors que le capital humain représenté par les scientifiques est tout au contraire extrêmement spécialisé: que l'on essaie de recycler un agronome en spécialiste d'économie de l'éducation ou seulement de l'environnement!

Cela s'accompagne de fortes variations de la demande politique sur le plan de la RSSP: qu'il s'agisse de "revirements" idéologiques et climatiques, de "changements d'orientation" politiques, d'instabilités d'attribution sur le plan de l'administration régulatrice ou résultant de contraintes budgétaires, ou bien de modes intellectuels et des fluctuations correspondantes des priorités de recherche – par exemple en ce qui concerne le développement économique, la réforme de l'enseignement, le développement urbanistique, le rapport entre milieu et marché de l'emploi – pour citer une liste des principaux thèmes des dernières décennies (Thurn et al., 1984, p. C 28). Ces fluctuations sont, dans chaque domaine de la RSSP (par exemple en sociologie de la criminalité), encore bien plus marquées qu'au niveau d'agglomération du financement de la recherche dans son ensemble et élèvent à la nième puissance, du côté de la demande, les éléments spéculatifs immanents à la science même, qui résultent (comme dans l'art ou la mode) de la création d'une pléthore de nuances stylistiques (dans la recherche par le moyen de concepts, de modèles, de séries de données etc.).

Les risques relatifs aux capacités et les fluctuations ont pour effet de rendre un brusque cumul de commandes ou une surexploitation soudaine des capacités tout autant problématique qu'une insuffisance de commandes ou un sous-emploi des capacités. Le programme de recherche allemand intitulé *Humanisierung der Arbeit* (humanisation du travail) peut servir d'exemple représentatif pour montrer combien trop d'argent peut

dépasser et démoraliser une communauté de chercheurs non préparée du point de vue professionnel et de ce fait ruiner, au lieu de la stimuler, une production compétente de savoir. Les risques relatifs aux capacités sont répartis de façon trés inégale: ils augmentent avec la taille de l'unité de recherche et avec les coûts fixes, avec le degré de spécialisation des chercheurs et donc avec la dépendance d'un petit nombre de grosses commandes ainsi qu'avec la partie des fonds de couverture qui ne proviennent pas de subventions et doivent être mobilisés sur le marché et qui souvent représentent déjà la plus grande partie des moyens financiers nécessaires.

Ces instabilités sont encore accrues par les segmentations existant à l'intérieur du domaine de la RSSP: segmentations entre la recherche universitaire et la recherche extra-universitaire; segmentations en fonction des spécialisations disciplinaires au sein des institutions, rendant très difficile le travail "interdisciplinaire" tant invoqué; segmentations selon les domaines établis de la RSSP (ainsi il n'existe par exemple aucun rapport entre la sociologie industrielle académique et l'étude de marché commerciale pour l'industrie et la politique); ou bien segmentations en fonction des relations établies avec la clientèle, relations qui protègent et favorisent une clientèle déterminée et évincent tout "outsider" sur les marchés oligopolistiques, ce qui vaut surtout pour les segments relativement stables des études de marché, d'opinion et de la conjoncture.

Les conséquences de ce que les économistes qualifieraient de "marché défectueux" sont évidentes: au cours des périodes marquées par une "pléthore" de commandes (d'Etat) la poussée des prix est remplacée par une baisse radicale du niveau de la qualité (baisse que la corporation doit alors renier en la définissant de critique de l'*overselling* et par conséquent comme attentes exagérées de la part des auteurs de la commande); en période de crise prédomine une compétition ruineuse (au sens propre du mot allant jusqu'à la destruction de l'existence bourgeoise). La compétition sur le plan de la qualité et des prix telle qu'elle est décrite dans les manuels d'économie est remplacée par une compétition, essentiellement politique et tendant à l'éviction mutuelle, sur le plan des fondations et des fermetures d'instituts.

La prédominance des composants politiques dans cette compétition est, non en dernier lieu, due aux particularités d'un métier qui ignore largement toute concession: à la différence par exemple de la branche voisine qu'est la psychologie, où des titres de légitimation sont parfois imposés par la loi, les chercheurs en sciences politiques et sociales n'ont pas besoin de licences. L'accès facile au marché, les besoins modestes de capitaux et un marché du travail tendu dans le domaine des sciences sociales font des crédits et des "relations" politiques (en tant que capital institutionnel et social) des ressources décisives. D'où résultent par la suite les inconvénients fâcheux tant déplorés que nous rencontrons tous les

jours: le dilettantisme de haut niveau des notables du demi-monde de la publicité scientifique, ces "Messieurs les professeurs *call-girls*" (A. Köstler) en tant qu'experts de tout, des agressions existant parmi les congénères jusqu'à la fin du monde en passant par la dignité de l'homme, y compris toutes les demandes spéciales et toutes les expertises de complaisance des clients – de préférence en termes "fédéraux allemands" ou "américains"; et enfin cette variante autrichienne propre aux sujets consistant à faire antichambre ("küss die Hand") chez les potentats et où les promoteurs renommés doivent être suivis jusqu'aux terrains de football, aux loges d'opéra ou aux chalets de montagne. Celui qui ne peut s'y faire est éliminé (ou émigré).

Tout cela ne fait que mettre en évidence une fois de plus le pouvoir écrasant détenu par certains organismes publics sur le plan de la demande, sans que les institutions étatiques et para-étatiques s'imposent pour autant (comme c'est le cas dans les branches caractérisées par un plus haut degré de professionnalisme) le régulateur de dispositions d'attribution coercitives pour elles-mêmes. C'est ainsi que par exemple en Autriche le principe de concurrence qui prescrit l'adjudication des marchés publics est pour ainsi dire complètement contourné dans le domaine de la RSSP. Il se peut en fin de compte que les produits et les prestations de la RSSP ne répondent, de par leur caractère, que de façon limitée aux nécessités du marché et des contrats: par nature vague et largement incertaine quant à son succès, comme c'est le cas pour toute recherche scientifique, la RSSP, si tant est que ce soit possible, est difficile à juger, et souvent seulement après coup; ce que "vaut" une enquête ou une simulation (pour l'auteur de la commande ou d'un point de vue "purement scientifique") ne se laisse déterminer de façon certaine que par la suite, parfois bien plus tard. En attendant, des contractants ignorants ou bien tout simplement habiles pourraient probablement contester la validité de la plupart de nos produits en raison d'une "réduction de plus de la moitié", si les spécialistes en sciences économiques et sociales commençaient à opérer sérieusement sur les marchés – ce qu'ils ne font pas. Comment expliquer autrement le fait qu'un seul et même type de prestation de recherche soit évalué parfois avec une différence de mille pour cent (!) et que l'absence de résultats entraîne néanmoins des projets consécutifs de plus grande envergure?

## Quel est le pouvoir du client?

Les relations entre l'auteur et l'exécutant d'une commande de RSSP sont, à deuxième vue, un peu plus difficiles à comprendre que ne laisserait supposer la conscience du rang professionnel. Bien entendu, le pouvoir dominateur des organismes publics sur le plan de la demande ne peut être mis en doute. Cependant il est loin d'être clair *comment* le pouvoir

concentré des auteurs des commandes, étant donné toutes les formes arbitraires d'attribution, s'assure réellement et dans chaque cas particulier une influence concrète. Tout aussi peu évidents me paraissent d'une part les déterminants du besoin de savoir arrêté en fonction de la politique, et d'autre part les constellations d'intérêts à la base. A ce propos voici quelques observations et thèses en vue d'une généralisation de l'évidence empirique disponible.

L'expansion rapide des instituts de recherche privés et d'utilité publique dans l'Europe de l'après-guerre est elle-même un premier indicateur du fait que les organismes publics ont éprouvé leur position quasi monopsoniste sur le plan de la demande tout autre que puissante: précisément les académies des sciences dépendant de l'Etat ainsi que les instituts de recherche universitaires et extra-universitaires étaient suffisamment "indépendants" pour frustrer de façon durable les besoins d'informations des gouvernements et des administrations. La sécurité absolue de l'emploi des scientifiques grâce au statut de fonctionnaires et le financement inconditionnel de la recherche à partir des budgets des collectivités territoriales protégeaient justement les instituts dépendant de l'Etat contre toute influence gouvernementale, ne serait-ce que sur le choix des thèmes. Paradoxalement ce n'est qu'avec la naissance d'une multitude d'instituts de recherche indépendants, échappant à toute possibilité de surveillance de la part des autorités, que s'est formé un pouvoir de la bureaucratie publique fondé sur l'adjudication des commandes et le financement des contrats. A la différence de la forme d'organisation juridique, une position de force des clients publics dans le domaine de la demande ne peut pratiquement naître que vis-à-vis d'instituts "indépendants", non subventionnés, non directement tenus de rendre officiellement des comptes et non soumis à une surveillance officielle directe (comme c'est le cas pour les universités). En conséquence l'administration publique participa et participe, souvent de sa propre initiative, à la fondation d'instituts autonomes qui – du point de vue des acheteurs d'informations – unissent les avantages d'une plus grande légitimité à ceux d'une plus grande accessibilité en ce qui concerne les désirs des clients.

La structure de la demande de RSSP entraîne, à travers les différentes formes juridiques, d'organisation et de financement une différenciation supplémentaire du champ en question: désormais on pourra réfléchir sur le chamanisme ou le concept du bonheur dans la philosophie sociale – chose possible jusqu'ici seulement dans les universités – et sur pratiquement n'importe quoi, de la motivation au travail jusqu'au milieu d'habitation, également en dehors des universités et bien entendu contre rémunération, à condition de tenir compte des besoins d'information de l'administration d'où émane la commande. Or, si l'on pose la question de savoir comment cela se passe exactement, on découvre deux caractéristiques dans les rapports entre auteurs et éxécutants des commandes dans le domaine de la

RSSP: premièrement les conditions d'un contrôle efficace de la production de savoir sont souvent également le préalable d'une augmentation des commandes; deuxièmement les auteurs des commandes détiennent, certes, tout le "pouvoir" (sur le plan de la demande), mais n'ont souvent que peu d'influence concrète sur les commandes. Alors que cela n'est guère contesté en tant que fait, les explications données à cet égard divergent.

Je ne partage pas les interprétations présomptueuses qui attribuent aux "bureaucrates" une incapacité fondamentale de commissionner la recherche, incapacité avancée par exemple pour expliquer l'échec de la mise en oeuvre des recommandations du *first Rothschild report* en Grande-Bretagne. (Cela s'applique peut-être davantage aux sciences naturelles). Entre-temps de nombreux ministères à travers l'Europe disposent de bonnes équipes qui, dans le cadre du processus de recherche, essayent d'intervenir, outre dans la simple adjudication des commandes, jusque dans le choix des méthodes, la formulation des questionnaires ou les méthodes d'enquête par sondage. C'est précisément l'existence de ce type de "bureaucrate engagé" (politiquement ou scientifiquement) et compétent dans sa branche, qui signale premièrement la différenciation interne des administrations publiques qui empêchent toute manipulation unidimensionnelle par des appareils monolithiques et qui souvent sont l'origine même de toute "demande" de recherche tendant à la critique de la société (il suffit de penser à la recherche marxiste sur le plan de l'urbanisme en France ou aux analyses portant sur le système de la santé publique, les classes et les crises en Autriche), ce qui fait qu'on réussit toujours à déceler, par exemple en Italie, un marché d'offres en matière de RSSP; et enfin le simple fait qu' "intervention" ne signifie pas forcément surveillance et que le dialogue et la communication peuvent parfois indiquer plus d'autonomie (reliée è une certaine politique sociale) que l'indifférence à l'égard de l'avancement de la recherche. Car c'est un client capable de diriger – de façon suffisamment efficace, par la formulation du thème, par les conditions d'attribution et surtout par le choix de l'institut ou du personnel – les résultats attendus, ou bien un client effectivement curieux de savoir à quoi aboutira la recherche, qui peut le plus facilement se permettre une attitude d'indifférence à l'égard du *processus* de recherche. (Il est typique que par exemple les associations des chefs d'entreprise, les syndicats ou les gouvernements s'adressent tous à des équipes de chercheurs différentes pour faire réaliser des simulations des effets d'une réduction de la durée du travail – l'exception étant l'Autriche.)

Cela nous amène à la question de savoir ce qui détermine au fond la demande politique relative à la RSSP et quels intérêts entrent en jeu dans ce contexte. Je pars de la supposition (inhabituelle) selon laquelle le besoin de savoir "objectif" n'augmente de façon spectaculaire que dans le sous-système que représente la science elle-même (plus on sait, plus on peut et plus on doit encore savoir), tandis que ce besoin ne change que très peu (et

plutôt de manière cyclique) dans la politique, étant donné que c'est le "demi-savoir" (*half-knowledge*) qui est nécessaire (Marin 1981). Le fait qu'une pression soi-disant objective exercée par un problème déterminé ne se transforme pas d'elle-même en un plus grand besoin de savoir, se manifeste lorsqu'on considère qu'on remarque actuellement une pression croissante exercée par les problèmes et en même temps une diminution des moyens d'exécution des commandes en RSSP. De toute évidence ce ne sont pas les objectifs déclarés tels la formulation de la politique et l'amélioration de l'information mais les fournitures de légitimation/consensus qui sont les déterminants politiques essentiels pour la demande de RSSP. Ce sont les fluctuations de ce besoin de consensus politique qui déterminent la demande sur le plan de la recherche en sciences sociales. On pourrait avancer la thèse que les gouvernements et régimes enclins aux réformes – *ou* instables et dans une situation politique fâcheuse – (comme par exemple actuellement en Scandinavie, en France, en Autriche, en Hongrie, en Italie, en Pologne et en Yougoslavie) ont bien plus "besoin" de la RSSP que les régimes à structure conservatrice *ou* les régimes incontestés (Etats-Unis, URSS, Grande-Bretagne, Bulgarie). Cela vaut également pour le niveau institutionnel à l'intérieur de chaque pays: les grandes associations à structure conservatrice *et* largement incontestées comme par exemple l'*OGB* (Confédération des travailleurs autrichiens); et la *Handelskammer* und *Arbeiterkammer* (Chambre de Commerce et Chambre des Travailleurs) en Autriche se permettent le soi-disant "luxe" d'un pronostic dans le domaine de la RSSP, toutes proportions gardées, bien moins souvent que des "outsiders" plus faibles tels l'*Industriellenvereinigung* (Association des industriels). Les différences, sur le plan de la demande relative à la RSSP, *entre* les pays s'expliquent probablement moins par des pénuries d'argent variant selon les caisses publiques que par des facteurs de ce qu'on appelle la culture politique: c'est ainsi que par exemple une bureaucratie administrative professionelle, qui survit à tous les changements de gouvernement (comme en Autriche), ne manifestera que peu d'intérêt pour la recherche évaluante. De cette manière on pourrait avancer et tester un certain nombre d'autres hypothèses. Mais, au lieu de cela, je voudrais approfondir brièvement la question, effleurée déjà plusieurs fois, concernant les fluctuations de la conjoncture politique de la RSSP et les problèmes d'adaptation des sciences sociales.

## Fluctuations de la conjoncture politique et problèmes d'adaptation des sciences sociales

Si la thèse selon laquelle le besoin de consensus est le principal déterminant de la demande de RSSP est valide, des modèles de développement politique divergents devraient également se traduire par des cycles de

demande divergents (voire par l'absence de ces cycles). On pourrait s'attendre à une demande politique accrue de RSSP pour les "nouvelles" politiques et les réformes qui, de façon chronique, manquent de légitimation et ont besoin de consensus, alors que les périodes et les organes de "restauration" peuvent plus facilement se passer d'une recherche acccessoire en sciences sociales. En effet la situation en Scandinavie, stable à relativement long terme, contraste par exemple avec les cycles de demande, caractérisés par une fébrilité à court terme, que provoquent les *crash programs* aux Etats-Unis ou en Europe occidentale: les programmes du gouvernment Johnson intitulés *War on Poverty* et *Great Society*, les programmes réformateurs introduits par la première coalition sociale-libérale en République Fédérale d'Allemagne, les premiers gouvernements de la Cinquième République en France ou les premiers gouvernments du centre gauche en Italie apportèrent sans exception un essor énorme de la RSSP, essor suivi non seulement d'une nouvelle contraction, mais également d'un cycle spécifique de RSSP demandée: consultation lors de la formulation de programmes, recherche portant sur l'action lors de la mise en application, recherche tendant à une évaluation en vue du contrôle des résultats, recherche visant la mise en oeuvre, une fois que la mise en pratique a été reconnue comme problème permanent (Thurn et al. 1984, p. C 23).

Sans aucun doute le penchant aux réductions fiscales, constaté au cours des phases de restauration qui suivirent, a-t-il affecté la RSSP bien plus profondément qu'on ne l'eût cru possible dans l'euphorie des années 60. De toute façon il ne fallait guère craindre une résistance décisive de la part d'un groupe si faible sur le plan social. En outre, ce ne fut la plupart du temps qu'à un stade avancé des développements politiques (lorsque quelque chose "s'était déjà passé") qu'on fit appel aux experts en sciences sociales, étant donné qu'on pouvait s'en passer temporairement. Enfin entrait également en ligne de compte le désavantage que représentait le "faible poids" financier qui – à la différence des grands projets technico-économiques – ne faisait pas craindre de "ruines d'investissement" visibles et écartait dès le début le syndrome toxicomaniaque caractérisé par le besoin de doses de subvention toujours plus fortes. Or, cette vue d'ensemble pessimiste nécessite elle aussi quelques importantes retouches que l'auto-apitoiement professionnel feint de ne pas voir.

Premièrement les gouvernments conservateurs n'ont pas réduit en général les dépenses de RSSP: la Suède et dans une certaine mesure Israel au cours des années 70 n'ont connu aucun "changement de cap" à cet égard, et le premier gouvernment Nixon, à tendance conservatrice plutôt modérée, ou le gouvernment Kohl/Genscher ont plutôt subventionné une *autre* RSSP qu'aucune. C'est ainsi que l'*Experimental Housing Allowance Program* sous Nixon a été par exemple le plus grand programme expérimental en matiére de politique et de RSSP jamais financé par des fonds publics

aux Etats-Unis. Mais même des gouvernements radicalement conservateurs comme ceux de Reagan ou de Thatcher, dont la politique à caractère populiste et quelque peu démagogique comporte de toute évidence des éléments anti-intellectuels et hostiles à l'éducation et aux sciences (sociales), se sont vite heurtés à certaines limites de leur *containment-strategy*. Hellmut Wollmann (in: Thurn et al. 1984, p. 115 et suivantes) a démontré pour les Etats-Unis que d'une part des restrictions budgétaires énergiques par exemple à l'égard de la *Division of Social and Economic Science* de la *National Science Foundation* (NSF)[5] ont dû être annullées au cours des années suivantes en raison de l'opposition inopinément forte de la part des scientifiques; que d'autre part des spécialistes en sciences sociales et économiques étaient eux aussi capables de mettre sur pied des organisations d'intérêts interdisciplinares (COSSA) et d'exercer par conséquent des activités de groupes de pression à la manière d'un lobby, et cela avec grand succès; qu'enfin les tentatives de *rollback*, étant donné la "masse critique" qu'a atteinte le nombre de spécialistes en sciences sociales et économiques aux Etats-Unis, ont abouti à un déplacement de l'activité et des subventions en matière de RSSP plutôt qu'à une réduction à cet égard: des programmes politiques nationaux ont été transmis à des Etats fédéraux sans être complètement dissous; des alliances avec des départements bienveillants de l'administration contournent souvent les directives du gouvernemnt fédéral; le Congrès, de par sa rivalité avec la Présidence, s'est fait le grand avocat de la RSSP; un grand nombre de fondations privées fait également fonction de "tampon financier" et compense souvent les restrictions de fonds publics destinés à la RSSP.

Voilà que nous venons d'aborder les conditions essentielles d'une capacité d'adaptation des sciences sociales à des constellations politiques variables sur le plan de la demande, que l'on pourrait décrire sommairement comme institutionnalisation. En font partie le devoir, imposé par la loi aux organismes publics, de se procurer des informations (devoir ayant abouti par exemple dans le *Civil Rights Act* de 1964 au célèbre *Coleman report*), et un réseau de fondations indépendantes, dont n'existent en Europe, hélas, que les premières ébauches[6] Aux Etats-Unis il existait déjà dans l'Entre-deux-guerres plusieurs milliers de fondations privées dont le nombre est désormais supérieur à 25.000 et dont le capital-actions dépasse de loin 25 milliards de dollars; les 12 plus grandes détiennent à elles seules un tiers de ces actions et subventionnent continuellement la RSSP (notamment dans les domaines des problèmes raciaux, de la criminalité et des relations internationales) (Horowitz/Katz 1975, p. 17 et suivantes). De tels stabilisateurs n'empêchent cependant pas que justement aux Etats-Unis on reste souvent attaché à un "modèle de sélection des marchés" qui se base sur quelque chose comme l'image naive d'une "crise purificatrice". Carol Weiss (WZB 1983) par exemple prétend qu'une trop forte demande politique de RSSP a abouti à trop de recherche ad hoc présentant peu de

qualité méthodologique et théorique, et compte sur la crise pour éliminer du marché justement les "unités de recherche marginales" de manière à ce qu'un nouvel essor garantisse à l'avenir une meilleure RSSP – comme si la compétition d'éviction ne se faisait également que de façon marginale selon des critères de qualités.

*Le Social Science Research Council* (SSRC) britannique par exemple a par contre pris la décision stratégique de principe de maintenir la RSSP dans les universités, de ne financer aucun centre de recherche indépendant et extra-universitaire, mais de promouvoir à la place de petits projets de recherche de la part de professeurs universitaires. Ce "modèle éponge" employé pour surmonter la crise part de l'hypothèse qu'il absorbera en quelque sorte les subventions – dans la mesure où elles existent – de la part d'institutions existantes sans que cela leur porte préjudice au cas où le financement complémentaire tarirait lors d'une pénurie de fonds. Ce *dual support system* – consistant en un financement de base et un financement complémentaire par des Conseils de recherche ou des fondations – serait en principe parfaitement valable, s'il n'était pas limité aux universités, car il faut se rendre à l'évidence que premièrement les instituts de recherche indépendants existent et que deuxièmement ils tirent leur raison d'être proprement du fait que les capacités de recherche des universités sont limitées. A cela s'ajoute souvent la divergence entre d'une part la recherche "valable" du point de vue académique (qui existe bien entendu également en dehors des universités) et d'autre part la RSSP intéressante du point de vue politico-pratique (qu'on trouve bien entendu également dans les universités): c'est ainsi que par exemple même une excellente recherche en matière de stratification et de mobilité serait sans aucune importance pour la politique sociale du moment que ses variables explicatives (comme par exemple le niveau de formation des parents) ne sont pas des paramètres influençables politiquement.

Un moyen de sortir de la cyclicité coûteuse et ruineuse de la RSSP pourrait éventuellement consister en une politique et une planification de recherche qui renversent l'actuelle stratégie de promotion et ne procèdent plus à des fondations d'instituts que lors des périodes de récession. Ainsi pourrait-on éviter le *paradoxe temps/argent*: lorsqu'il y a une demande de RSSP et que l'argent est disponible, il manque le temps pour effectuer le travail de base nécessaire; lorsque le temps y est, il manque l'argent pour utiliser qualitativement les surcapacités créees. Une politique de fondation en périodes de pénurie d'argent aigue permettrait par contre dès le début une concentration sur les problèmes fondamentaux des méthodes, la construction de modèles et d'autres activités créatives, le capital humain accumulé pouvant être employé de manière d'autant plus productive au cours des phases d'expansion successives.

Je ne peux pas ici pousser plus loin ces réflexions, or la question se pose tout de suite de savoir si cela ne dépasse pas les horizons temporels de

l'actuelle politique en matiére de science. Dans la pratique semble plutôt dominer une sorte de *recherche à souffle court "à la manière des sapeurs-pompiers"*, que je voudrais illustrer par un seul exemple. Le lendemain du commencement, en 1981, de l'agitation des jeunes, qui se propageait de Zurich à Vienne, un institut de recherche indépendant fut chargé d'analyser "l'échauffourée". L'étude ne dépassa cependant pas la phase pilote – qui dura quelques mois – tout simplement parce que la rapide accalmie qui suivit "l'échaffourée" eut également pour effet de supprimer l'intérêt que les responsables de la fondation et les organismes publics avaient manifesté à l'égard de la compréhension du phénomène. Dit avec un léger cynisme, n'aurait-il pas été raisonnable que sciences sociales et vitrerie s'unissent pour "offrir" quelques glaces de vitrine aux enfants nous permettant ainsi d'étudier à fond, en fin de compte, leur morosité cachée?

Or, les fonds et les organismes de promotion qui ne subventionnent pas de recherche sur commande, mais seulement une recherche sur demande, ne sont, eux non plus, nullement à l'abri de brusques fluctuations politiques, de modes scientifiques, de l'humeur des conseillers et de changements d'interprétation des directives d'un programme à moyen terme – je renonce à citer des exemples à ce propos. Par contre je voudrais très brièvement exposer comment la situation décrite ci-dessus se répercute sur les publications qui furent naguère un autorégulateur central de la production du savoir scientifique.

## Problèmes subséquents illustrés en prenant pour exemple les publications

L'autoréflexion scientifique basée exclusivement sur les sciences (Dobrov 1969) a mis au premier plan "l'explosion des publications", déterminée dans sa croissance par des fonctions exponentielles logarithmiques, et ses déterminants, phénomène qui apparaît sans doute également dans le domaine de la RSSP. Il en résulte *entre autres* que la "période moyenne d'actualité" des articles en matière de sciences sociales s'est entre-temps réduite à 5 ans: ce qui se lit après devient quasi une contribution "classique" – ou le fait de lecteurs attardés. (Dans les sciences physiques et naturelles la "vitesse de circulation" est bien plus élevée). Même les lecteurs les plus zélés ne peuvent tout simplement pas venir à bout de l'avalanche de publications. L'introspection critique (et les études des citations) montre cependant qu'on ne lit pas davantage de ce fait ou en général, mais qu'on lit plutôt toujours moins – on ne lit même pas régulièrement les revues pour lesquelles on écrit (ou voudrait écrire). Le profond ennui qu'exhalent la plupart des revues qui portent sur une discipline déterminée ne me semble être qu'un prétexte à cet égard; ce phénomène signale plutôt la dégradation de la branche traditionnelle des

publications où la *peer review* était conçue comme un mécanisme de contrôle intraprofessionnel central. La communication et le contrôle sont respoussés au profit de revendications d'autorité sociale dans le domaine scientifique: celui dont les écrits sont (ou ont été) publiés ici et là, peut se mettre en avant, alors qu'une publication par une maison d'édition de moindre renommée peut signifier publication à l'écart du public (compétent en la matiére) ("Du bon travail, dommage qu'il ait paru là, c'est comme s'il n'avait jamais été fait.") Mais où cela mène-t-il (sinon au Japon), si un travail est presque toujours seulement aussi bon que l'institut d'où il provient et la revue dans laquelle il a paru?

Là-dessous se cachent, du moins dans les sciences sociales, de nombreux autres problèmes. Si les publications naissent pour ainsi dire comme "sous-produits" des revendications de statut en rapport avec la discipline, revendications que l'on peut également faire valoir d'une autre façon (par exemple par des expertises, des consultations, des commandes de recherches à grande échelle, un travail en comité etc.), alors il ne faut pas non plus s'étonner d'une "émigration des publications". Par là je ne vise pas la littérature "grise" qu'on n'arrive plus à embrasser dans son ensemble mais qui est devenue toujours plus importante – ce nombre infini de "séries", de "documents de travail", de *mémoranda*, de *discussion papers* ou de *working papers* – littérature qui anime et enrichit et parfois dirige en effet le flux des communications à l'intérieur des sciences, non je pense plutôt à ces formes de publication qui désertent les sciences mêmes tout en les influençant efficacement de l'extérieur: ce sont, d'un côté de l'échelle, les rapports de recherche "confidentiels" remis au client, dont les données et les résultats restent inaccessibles même à la "communauté des scientifiques", et de l'autre les publications dans des revues, quotidiens etc. non professionnels, de caractère "journalistique".

Tous deux ont de nombreuses causes et conséquences dont très peu seulement ont déjà été analysées, ce qui fait que je peux uniquement soulever quelques questions sans offrir de réponses. Or, il y a lieu de poser les questions et de chercher les réponses. Quelle est la signification, pour l'auto-orientation et le développement cumulatif des savoir des sciences sociales, du fait qu'une partie de leurs données et constatations – par exemple dans le domaine du sondage d'opinion politique – soit systématiquement écartée du processus de traitement théorique? Que signifient conservation du secret et droit de propriété sur le savoir (informations, banques de données, modèles, etc.; dans d'autres disciplines également licences, brevets, etc.) dans le système social des sciences qui revendiquait par tradition la publicité et le "communisme" du flux d'informations? Et qu'en résulte-t-il pour la démocratie si les citoyens n'ont plus accès justement aux savoir qui sont systématiquement recueillies sur leur compte?

Dans des espaces tout aussi peu explorés nous conduit le phénoméne relativement nouveau d'un journalisme des sciences sociales qui présente

lui-même plusieurs variantes. Or, on ne peut nier ce qui suit: les économistes les plus influents (dont certains lauréats du prix Nobel) écrivent régulièrement des articles pour des hebdomadaires afin de transmettre des messages simples à un vaste public; d'importants débats politiques (portant par exemple sur la *new class*, l'"*affermative action*", le sens et non-sens des tests relatifs au quotient intellectuel) sont engagés dans des centaines de revues non techniques telles *Public Interest*, *Commentary* etc.; les résumés de recherche se lisent parfois le mieux dans le *New Society* (en Grande-Bretagne, le *Spektrum der Wissenschaft* en RFA) etc.; certains des meilleurs spécialistes italiens en sciences économiques et sociales écrivent plus souvent pour le quotidien *La Repubblica* que pour des revues professionelles nationales; d'importantes analyses politiques se trouvent d'abord (ou seulement) dans *Die Zeit*, *Le Monde*, *Le Monde Diplomatique*, *Preuves*, *Contrepoint*, *Polytika*, etc. Que ces activités résultent plutôt des prétentions traditionnelles des mandarins, comme en France ou en Pologne, ou d'une combinaison de mauvaise rémunération et d'esprit d'entreprise, comme en Italie, ou de la naissance d'une nouvelle profession d'intermédiaire du journalisme scientifique, comme surtout en Grande-Bretagne et aux Etats-Unis, ou bien de tout autre chose – les conséquences à plus long terme de cette évolution qui reflète également le morcellement croissant du champ de la RSSP ne sont, selon toute apparence, pas prévisibles.

Une meilleure évaluation de ce phénomène constituerait une aide d'orientation extrêmement importante pour notre identité et notre travail en tant que spécialistes en sciences sociales. L'écarter purement et simplement en prétendant qu'il s'agit du zèle "extraprofessionnel" de quelques confrères affairés, qui reste sans conséquences pour notre recherche scientifique, serait par trop naif. Y voir l'érosion et la dégradation du modèle des sciences qui nous est familier et que nous tenons en grande estime du point de vue normatif, pourrait être justifié, sans pour autant nous être utile. Peut-être les tendances esquissées de ce développement signalent-elles simplement la naissance de nouvelles professions intellectuelles, cristallisées autour de nouveaux "créneaux du marché" de la production et de la transmission de savoir, auxquelles notre "métier de travailleur intellectuel" qu'il ne faut pas non plus trop vénérer – et qui se croit aux premiers rangs de la production de savoir en matière de sciences sociales – ne songe même pas.

## Bibliographie

Andrews, F.M. (Ed.), Organization and Performance of Research Units: A Six Country Study, N.Y. 1978.

Arculeo, A. et al., La scienza politica in Italia: materiali per un bilancio, Milano 1984.

Aron, J.P., Qu'est-ce que la culture française, Paris 1975.

Badura, B. (Hg.), Seminar: Angewandte Sozialforschung. Frankfurt/M 1976.

Baehr, P.R., Wittrock B. (eds.), Policy Analysis and Policy Innovation: Patterns, Problems and Potentials, Beverly Hills 1981.
Ben-David, J., The Scientist's Role in Society. A Comparative Study, Englewood Cliffs 1971.
Beyme, K. von, Politische Kybernetik? Politik und wissenschaftliche Information der Politiker in modernen Industriegesellschaften, in: Journal für Sozialforschung 24 (1984), 1, pp. 3-16.
– La scienza politica in Italia: uno sguardo dall'estero, mimeo, Milano 1984.
– Policy Analysis und traditionelle Politikwissenschaft, mimeo, Heidelberg, 1984.
Bobbio, N. Saggi sulla scienza politica in italia, bari 1971.
Boudon, R., Effets pervers.
Bourdieu, P. Champs du pouvoir, champs intellectuel et habitus de classe, in: Scolies, 1, 1971, pp. 7-26.
– Boltanski, L., Malidier, P., La defense du corps, in: Informations sur les sciences sociales X (1971), 4.
– Boltanski, L., De Saint Martin, M., Les Strategies de reconversion, in: Informations sur les sciences sociales, XII (1973),5
– Le champs scientifique, in: Actes de la recherche en sciences sociales, 2-3, 1976, pp. 88-104.
– La production de la croyance, contribution a une economie de biens symboliques, in: Actes de la recherche en sciences sociales, 13, 1977, pp. 3-40.
– Classement, déclassement, reclassement, in: Actes de la recherche en sciences sociales, 24, 1978, pp. 2-22.
– Questions de Sociologie, Paris, 1980.
– Homo Academicus, Paris 1984.
Broad, W., Wade N., Betrayers of the Truth, N.Y. 1982
Bulmer, M. (ed.), Social Science Research and Government, Cambridge, 1986.
Caplow, T. Mc Gee, R.J., The Academic Marketplace, N.Y. 1965.
Chamboredon J.C., Sociologie de la sociologie et interets sociaux des sociologues, in: Actes de la recherche en sciences sociales I (1975), 2, p. 12 f.
Clark, T.N. Prophets and Patrons, the French University and the emergence of social sciences, Cambridge, Mass. 1973.
Coleman, J.S., Conflicts Between Policy Research and Decision Making, pp. 14-29 in H. Skoje (ed.), Scientific Expertise and the Public, Oslo, 1979.
– The Structures of Society and the Nature of Social Research, in Knowledge, 1, 1979 pp. 333-50.
– J. Policy Research in the Social Sciences, Morristown 1972.
Coser, L.A. Men of Ideas, New York 1965.
Crane, D., Invisible Colleges, London 1972.
Crawford, E, Mazoyer, M., Barthelemy, C., Enquete Europeenne sur l'organisation de la recherche en sciences sociales, CNRS, Paris 1973.
– Eléments d'une evaluation des centres de recherche en sciences sociales, CNRS, Paris 1976.
Cresal (eds.), Situations d'expertise et socialisation des savoirs, St. Etienne, 1985.
Daedalus, Science and Its Public: The Changing Relationship., Vol, 103 (1974), No.3
Dierkes, M. et. al. (eds.), Comparative Policy Research: Learning from Experience, Aldershat, 1986.
Dobrov, G.M., Wissenschaftwissenschaft, Berlin 1969
Dror, Y. General Policy Science, in: Nagel, S. (ed.), Policy Studies in America and Elsewhere, London pp. 3-16.
Elias, N., Problems of Involvement and Detachment, in: Brit. J. Soc. 7 (1956), pp. 236-254.

Fabris, H. et al, Zum Politologenbedarf in Österreich, in: ÖZP 4 (1973), S. 419 f.
Ferraresi, F., Ferrari, G., Italy: Economists in a weak political system, in: A.W. Coats (Ed.), Economists in Government, Durham, N.C. 1981.
Firnberg, H., Zur Rolle der Sozialwissenschaften in der österreichischen Wissenschaftspolitik: das Anwendungsdefizit der Soziologie, in ÖZS 3 (1978), 1, pp. 4-10.
Firpo, L, La facoltà di scienze politiche, in: Il politico, 1967, pp. 667-687.
Fischer-Kowalski, Zur Entwicklung von Universität und Gesellschaft in Österreich, in: Fischer, H. (Hrsg.) Das politische System Österreichs, Wien 1974, pp. 571-624.
Foucault, M. Power/Knowledge, New York 1980.
Fraisse, R., Les sciences sociales: utilisation, dependance, autonomie, in: Sociologie du travail, Vol. 23 (1981), No. 4.
Friedberg, E., Gremion, Pl, La recherche administrative et le réformisme politique, Paris, CSO, mimeo, 1974.
Giacomoni, S., Miseria e nobiltà della ricerca in Italia, Milano 1979.
Geiger, Th., Aufgaben und Stellung der Intelligenz in der Gesellschaft, Stuttgart 1949.
Giddens, A., Central Problems in Social Theory, London 1979.
– The Constitution of Society, Cambridge 1984.
Glaser, B.G., Organizational Scientist, Indianapolis, 1964.
Goffman, E., On Cooling the Mark Out: Some Adaptation to Failure, in: Psychiatry 8 (1952), pp. 451-463.
Gouldner, A.W. (ed.), Cosmopolitans and Locals: Toward an Analysis of Latent Social Roles, in: Adm. Sc. Qu. II (1957/58), pp. 281-306 und VIII (1963), pp. 373-380.
– The Coming Crisis of Western Sociology, London 1971.
Habermas, J., Technik und Wissenschaft als "Ideologie", 2. Aufl. Frankfrut/M. 1969 (1968).
Hammond, Ph.E., Sociologist at Work: the Craft of Social Research, N.Y. 1964.
Hartmann, H., Organisation der Sozialforschung, Opladen 1971.
Hausen, K. Nowotny, H. (Hrsg.), Wie männlich ist die Wissenschaft?, Frankfurt/M. 1986.
Heidenheimer, A.J., Politics, Policy and Police as Concepts in the Western Languages or: Why are the "Kontis" deprived?, mimeo Chicago 1983 (APSA).
Hellstern, G.-M., Wollmann, H. (Hrsg.), Experimentelle Politik, Opladen 1983.
Hetzler, H. W., Soziale Strukturen der Organisierten Forschung. Düsseldorf 1970.
Holt, R.T., Tuner, J.E., The Scholar as artist in: Policy Science 1974, pp. 257-270.
Horowitz, I.L. et al,, Social Science and Public Policy in the United States, N. Y. 1975.
Jann, W., Der Policy-Ansatz. Speyerer Arbeitshefte 45, 1983
Jay, M., The Dialectical Imagination. A History of the Frankfurt School and the Institute of Social Research, London 1973.
Knorr, K., et al, Sozialwissenschaftliche Forschung in Österreich, Wien 1981.
Konrad, G., Szeleny, I., Die Intelligenz auf dem Weg zur Klassenmacht, Frankfurt/M. 1981.
Köstler, Die Herren callgirls
Kuhn, T.S., The Essential Tension: Tradition and Innovation in Scientific Research, in: Hudson, L. (ed.) The Ecology of Human Intelligence, Harmondsworth 1970, pp. 342-356.
Lasswell, H.D., The Policy Orientation in: Lerner, D. Lasswell, H. D. (eds.) The Policy Sciences, Stanford 1951. pp. 3ff.
Lazarsfeld, P.F., The Sociology of Empirical Social Research, in: ders. 1972, pp. 321-340.
– Merton, R. K., A Professional School for Training in Social Research, in: ders. 1972, pp. 361-391.
– Qualitative Analysis. Historical and Critical Essays, Boston 1972.
– Eine Episode in der Geschichte der empirischen Sozialforschung, in: Parsons, T., et al., 1975, pp. 147-225.
Lentini, O., La ricerca sociale in Italia tra prima e seconda fondazione della sociologia in: Sociologia, Vol. 12 (1978) No. 3.

Leoni, B., Un bilancio lamentevole: il sottosviluppo della scienza politica in Italia, in: Il Politico, 1960, pp. 31-42.
Lepenies, W. (Hg.) Geschichte der Soziologie, Frankfurt/M. 1981, 4 Bde.
Die Drei Kulturen. Soziologie zwischen Literatur und Wissenschaft, München 1985.
Lepsius, R.M. (Hg.) Soziologie in Deutschland und Österreich 1918-1945, Opladen 1981.
Levine, R. A., et al. Evaluation Research and Practice, Beverly Hills, 1981.
Lindberg, L. Wirtschaftswissenschafter als Politikberater, in: Journal für Sozialforschung, 23 (1983) 1, p. 3f, 2, pp 25,2, 185-204.
Lindblom, C. Cohen, D.K., Usable Knowledge. New Haven 1979.
Lompe, K. et al, Enquete-Kommissionen und Royal Commissions, Göttingen, 1981.
Luhmann, N., Theoretische und praktische Probleme der anwendungsbezogenen Sozialwissenschaften: Zur Einführung, in: Wissenschaftszentrum Berlin (Hrsg.) 1977.
Lutz, B., Zur Lage der soziologischen Forschung in der Bundesrepublik – Ergebnisse einer Enquête der Deutschen Gesellschaft für Soziologie, in: Soziologie, 1977, Heft 1, pp. 4ff.
Machlup, F., The Productivity and Distribution of Knowledge, Princeton 1962.
Mannheimer, R., Repertorio delle ricerche empiriche e dei files in Italia in: Arculeo et al. 1984.
Marin, B., Politische Organisation sozialwissenschaftlicher Forschungsarbeit, Wien 1978.
– What is 'Half-Knowledge' Sufficient For – and When? Theoretical Comment on Policymakers' Uses of Social Science, in: Knowledge: Creation, Diffusion, Utilization, Vol. 3 (1981) No.1, pp. 43-60.
– Der Beirat für Wirtschafts- und Sozialfragen, in: ders., Die Paritätische Kommission, Wien 1982, pp 255-410.
– The Development and State of Policy Research in Austria, mimeo Berlin 1984.
– Generalized Political Exchange, Preliminary Considerations, EUI working Paper No. 85/190, Florence 1985.
Max-Planck-Gesellschaft, Die MPG und ihre Institute, München 1972.
Mayntz, R., Forschungsmanagement – Steuerungsversuche zwischen Scylla und Charybdis, Opladen 1985.
– The Conditions of Effective Public Policy: a New Challenge For Policy Analysis, in: Policy and Politics, Vol. 11 (1983), No. 2, pp. 123-143.
– Struktur und Leistung von Beratungsgremien: Ein Beitrag zur Kontingenztheorie der Organisation, in: Soziale Welt, (1977), Heft, 1-2 pp. ...
Meltsner, A. J. Policy Analysts in the Bureaucracy, Berkeley, 1976.
Merton, R.K., The Ambivalence of Scientists, in: Bulletin of the John Hopkins Hospital, 1963, pp. 77-97.
– The Matthew Effect in Science, in: Science 159 (1968), pp. 56-63.
Mills, C.W., On Intellectual Craftmanship, in: ders. The Sociological Imagination. N.Y. 1961 (1959), pp. 195-226.
Mosher, F.C.,The GAO: The Quest for Accountability in American Government, Boulder, CA 1979.
Mulkay, M.M., The Social Process of Innovation, London 1972.
Müller-Rommel F., Sozialwissenschaftliche Politikberatung, in: Politik und Zeitgeschichte B 25/84 (1984), pp. 26-39.
Nagel, S., Policy Studies and the Social Sciences, Lexington, 1979.
Nowotny, H., Lambiri-Dimaki (eds.), The Difficult Dialogue between Producers and Users of Social Science Research, Vienna 1986.
Offe, C., Sozialwissenschaften zwischen Auftragsforschung und sozialer Bewegung in: Greiff, B. (Hg.), Das Orwell'sche Jahrzehnt und die Zukunft der Wissenschaft, Opladen 1981, pp. 98-108.

O'Toole, R. (ed.) The Organization, Management and Tactics of Social Research, Cambridge, Mass. 1971.
Parsons, T., et al., Soziologie- autobiographisch, Stuttgart 1975.
Perry, N., Crawford, E., Demands for Social Knowledge, Beverly Hills-London, 1976.
Pichieri, A., Policy Research in Italy, mimeo Torino 1983.
Pinder, J. (ed.) Fifty Years of Political and Economic Planning, London 1981.
Pinto, D., La Sociologie dans l'Italie de l'apres-guerre, 1950-1980,in: Revue francaise de sociologie, Vol. 21 (1980) No.2, pp. 233-250. Pinder, J. (ed.), Fifty Years of Political and Economic Planning, London 1981.
Polanyi, M. Personal Knowledge, N.Y. 1962.
Pollak, M., la planification des siences sociales, in: Actes de la recherche en scienes sociales, 2-3, 1976, pp. 105-121.
– Gesellschaft und Soziologie in Frankreich, Königstein/Ts. 1978.
Paul F. Lazarsfeld – Gründer eines multinationalen Wissenschaftskonzerns, in: Wolf Lepenies (Hg.) Geschichte der Soziologie, Frankfurt/M. 1981.
Rein, M., Social Science and Public Policy, Harmondsworth, 1976.
Ringer, F. K., The Decline of German Mandarins, The German Academic Community, 1890-1933, Cambridge, Mass. 1969.
Rose, R., Policy Research and Government Policy, Studies in Public Policy No. 100, Glasgow 1982.
Rosenmayr, L., Austrian Sociology – Past and Present. Notes on Some Historical Trends and Present Research Activities, in: Soc. Sc. Inf. V (1965), pp. 58-101.
– Zur Praxisrelevanz soziologischer Forschung, in: ÖZS 3 (1978), 1 pp. 11-18.
Lord Rothschild An Enquiry into the Social Science Research Council, London 1982.
Sartori, G. La scienza politica, in: Il politico, 1967, pp. 688-701.
Scamozzi, S., Lo studio della società: orientamenti storici, in: Paolo Farneti (ed.), Politica e società, Firenze 1979, Vol. 2. pp. 905-929.
Schneider, J.A., et al , Policy Research and Analysis: an Emprical Profile, 1975-1980, in Policy Sciences (1982), pp. 9ff.
Smith, B.L.R., The RAND Corporation, Cambridge, Mass. 1966.
Spreafico, A. Studi politici e scienza politica in italia, in: Annuario politico italiano, Milano 1964, pp. 202-230.
Stonequist,. E., The Marginal Man, N.Y. 1937.
Taylor, C.W. Barron, F. (eds.) Scientific Creativity: ITs Recognition and Development, New York 1975.
The New York Times Magazine, Hot Colleges. And How They Get That Way, 18, Nov. 1984, pp. 69-73.
Thurn, G. et al, Development and Present State of Policy Research, Wissenschaftszentrum Berlin, mimeo, 1984.
Torrance, J., The Emergence of Sociology in Austria 1885 – 1935, in: Arch. europ. soc. XVII (1976), pp. 185-219.
Vito, F. Gli studi politici in Italia, Milano 1964.
Vollmer, H.M., Mills, D.L., (eds.), Professionalization, Englewood Cliffs 1966
Wagner, P., Wollmann, H., Patterns of the Engagement of Social Scientists in Policy Research and Policy-Consulting. Some Crossnational Considerations, Berlin: WZB papers, p. 86-3, 1986.
Wagner, P., Social Sciences and Political Projects: The Emergence and Demise of Reform Coalitions between Social Scientists and Policy-Makers in France, Italy, and West Germany, forthcoming in: Stuart Blume et al. (eds) The Social Direction of the Public Sciences. Sociology of the Sciences Yearbook XI, Dordrecht, Reidel 1987.

Weber, M. Wissenschaft als Beruf, in: ders., Gesammelte Aufsätze zur Wissenschaftslehre, Tübingen 1951, pp. 524-613.
Weingart, P. Winterhager, M., Vermessung der Forschung, Frankfurt/M. 1984.
Weiss, C. Evaluation Research, 1972.
Wildavsky, A., Speaking Truth to Power: the Art and Craft of Policy Analysis, Boston 1979.
Wilensky, H.L., Organizational Intelligence, N.Y. 1966.
Wissenschaftszentrum Berlin (Hg.), Interaktion von Wissenschaftund Politik, Frankfurt/M. 1977.
– Development and Present State of Policy Research, minutes of the round-table discussion, Berlin June 1983.
Wittrock, B., Elzinga, Al, The University Research System: The Public Policies of the Home of Scientists, Stockholm, 1985.
– Social Knowledge and Public Policy: Eight Models of Interaction, in Nowotny, H., Lambiri-Dimaki, J. (eds.), 1985, pp. 89-109.
Znaniecki, F., The Social Role of the Man of Knowledge, N.Y. 1968.

## Notes

[1] Ainsi y avait-il déjà au début des années 70 en France, à cet égard loin d'être à l'avant-garde, plus de chercheurs sur des postes temporaires que sur des postes permanents à l'intérieur *et* à l'extérieur des universités, Crawford et al. 1973.

[2] Pour l'Autriche cf. entre autres l'exposé présenté par Bodzenta en décembre 1983 lors de l'assemblée annuelle de la Société Autrichienne de Sociologie à Klagenfurt; pour la France cf. Bourdieu, Boltanski, Ladidier, 1971.

[3] L'économiste et conseiller syndical Ezio Tarantelli, assassiné le 27 mars 1985 par les *Brigate Rosse*, a par exemple travaillé à la *Banca d'Italia* avant d'enseigner et tout au long de ses engagements au *Massachusetts Institute of Technology (M.I.T.)*, à l'*Institut Universitaire Européen de Florence* et à l'université de Rome.

[4] Récemment entre autres les universités *Duke*, *Brown* et de *Virginia* à côté des traditionnels *Yvy League Colleges*; voir The New York Times Magazine du 18 Nov. 1984.

[5] Le budget des sciences sociales et économiques, dans le cadre de la *National Science Foundation*, a par exemple été ramené de 31,3 millions de dollars en 1980 à 17,6 millions de dollars en 1982.

[6] Néanmoins des initiatives privées jouent parfois également en Europe un rôle important. En Italie par exemple, la recherche dans le domaine de la sociologie industrielle n'a été possible, avant la reconnaissance académique et universitaire de ce métier, que grâce aux subventions accordées par l'industrie; la même chose vaut pour l'aménagement urbain. Ainsi le groupe *Olivetti* de Turin a-t-il créé un institut de recherche à *Ivrea*, qui a élaboré le premier plan de développement de cette ville industrielle. (Alessandro Pizzorno qui enseigne actuellement à *Harvard* et qui est peut-être le spécialiste italien en sciences sociales le plus renommé de nos jours, a été, de 1953 à 1957, directeur du département de recherches du groupe *Olivetti* sur les relations du travail dans l'industrie.) Face à un Etat central faible et retardataire la bourgeoisie industrielle du nord de l'Italie, notamment au Piémont et en Lombardie, était une force progressiste et motrice du renouvellement social – également en ce qui concerne le développement des sciences économiques et sociales dans une culture universitaire dominée par la pensée philosophico-juridique.

# Der Souveränitätsbegriff in der Griechischen Verfassung 1975

ATHANASIOS MOULAKIS
San Diego/Florenz

Das Thema dieses Aufsatzes ergibt sich aus der Hervorhebung der Volkssouveränität in der griechischen Verfassung von 1975[1]. Diese Hervorhebung besteht, erstens in der Verdrängung des in der griechischen Verfassungstradition eingebürgerten Souveränitätsträgers – der "Nation" – durch das "Volk" und zweitens, im Aufrücken der so veränderten Souveränitätsformel von der bescheideneren Stelle in Artikel 21 der Verfassung von 1952 zum deklaratorischen Auftakt der Verfassung von 1975, zum Artikel 1.

Die traditionelle und noch im Verfassungsentwurf, den die Regierung 1975 einbrachte, zunächst enthaltene Formulierung hatte folgenden Wortlaut:

"Alle Gewalten gehen von der Nation aus, bestehen um ihretwillen und werden gemäß den Verfassungsvorschriften ausgeübt"[2].

Nach einer außerordentlich lebhaften Debatte verabschiedete das mit der Erstellung des neuen Verfassungstextes beauftragte Parlament folgende endgültige Fassung:

"Fundament der Verfassungsordnung ist die Volkssouveränität. Alle Gewalten gehen vom Volk aus, bestehen um des Volkes und der Nation willen und werden gemäß den Verfassungsvorschriften ausgeübt"[3].

Was hier vorliegt ist eine Veränderung des Stellenwertes politischer Symbole – Souveränität – Volk – Nation – anhand deren sich die Gesellschaft – oder zumindest ihre Repräsentanten im Parlament – über den Sinn gemeinschaftlichen Zusammenlebens, wenn nicht einigen, so doch klar zu werden versuchen.

Die Souveränitätsformel wurde mit großer Intensität um nicht zu sagen Heftigkeit in der Kammer und in der mit der Redaktion des Vertragsdokuments beauftragten Kommission diskutiert. Man hat sehr viel Zeit und Energie in der Behandlung dieser Frage aufgewandt, wie in der Diskussion grundlegender institutioneller Dispositionen – wie etwa die (neuerdings von der jetzt regierenden Mehrheit abermals veränderter)

Absteckung der Befugnisse des Staatsoberhaupts – was für die große Bedeutung spricht, die die Abgeordneten der ideologischen und emotionellen Kraft des Symbols beigemessen haben[4].

Meine These ist

1) Daß die in der griechischen Verfassungsdebatte von 1975 zum tragen gekommene Symbolsprache die Gefahr der doktrinären Deduktion aus ideologisch gesetzten Prinzipien in die Gestaltung von politischen Institutionen hineinträgt, und daß

2) das so gegebene Vokabular der adäquaten Artikulation mancher Repräsentationsprobleme hinderlich ist.

Andererseits möchte ich zeigen,

1) daß die historisch bedingte polemische Abgrenzung des neuen, sich eben konstituierenden Regimes gegen die 1974 gestürzte Diktatur die Anziehungskraft der Formel von der Volkssouveränität psychologisch erklärt.

2) Daß das verfassungsrechtliche Vakuum von 1974 die Annahme einer vorkonstitutionellen und immerwährenden souveränen Instanz aus recht logischen Gründen dienlich erscheinen ließ.

In diesem Zusammenhang werden Schwierigkeiten sichtbar, die sich für eine politische Problematik ergeben, wenn sie durch die Vorherrschaft rechtsformalen Denkens überlagert wird. Schließlich möchte ich zeigen,

3) daß innerhalb der terminologischen Schranken der Verfassungsdiskussion, trotz aller Probleme, historische Belastungen griechischen Lebens, sowohl im Bezug auf den inneren Frieden, als auch im Verhältnis zum Auslandsgriechentum abgebaut werden konnten.

An diesem Punkt ist es dienlich, einige historische Daten in Erinnerung zu bringen.

Eine Konstellation innen- und außenpolitischer Faktoren führte im Sommer 1974 zur Abschaffung einer siebenjährigen Militärdiktatur. Der Putsch von November 1973 hatte die – wie es sich erwies – letzte Phase der Diktatur eingeleitet. Unter der de facto einzig maßgeblichen und kaum verhüllten Befehlsgewalt des Chefs der Militärpolizei wurde eine denkbar inkompetente Führung der inneren wie der äußeren Angelegenheiten des Landes von einer Intensivierung der repressiven Maßnahmen des Regimes begleitet. Die Entdeckung von Öl in der Nähe der Insel Thasos führte zu Gegensätzen mit der Türkei über die respektiven Rechte in der Ägeis, ein Streit, der im Frühjahr 1974 den Krieg zwischen beiden Ländern nicht ausgeschlossen erscheinen ließ. Von Athen aus geführte Intrigen mündeten in einem Putsch auf Zypern und einem Attentat auf das Leben des Präsidenten Makarios. Dies veranlaßte – nach einigen diplomatischen Demarchen – die türkische Landung auf der Insel am 20. Juli. Daraufhin wurden sowohl die griechischen wie die türkischen Streitkräfte mobilisiert. Die Zypernkrise legte die internationale Isolierung des Athener Militärregimes bloß. Der gekränkte Nationalstolz der zivilen Bevölkerung

nicht weniger als der der Armee kam zu den bereits angedeuteten inneren Schwächen und Spannungen des Regimes hinzu. Die Moral und Effizienz der Streitkräfte in ihrer Hauptaufgabe, Krieg gegen äußere Feinde führen zu können, schien vielen aktiven Offizieren durch ihre politische Verstikung beeinträchtigt. Aber nicht nur die Prioritäten von Offizieren verwandelten sich vor der türkischen Drohung, sondern auch die Zusammensetzung der Armee aufgrund der in der Generalmobilmachung einberufenen Reservisten. Drei Tage nach der Türkischen Invasion verlangte der Befehlshaber der mächtigsten Armeeinheit des Landes, des im Norden stationierten III. Armeekorps, ultimativ die Wiederherstellung einer zivilen Regierung. Der Staatspräsident rief daraufhin eine Konferenz führender Militärs und prominenter Politiker aus der Zeit vor der Diktatur zusammen, von der der Ruf an K. Karamanlis ausging, aus seinem selbstauferlegten Pariser Exil zurückzukehren und die Leitung der Regierungsgeschäfte zu übernehmen. Schon in der Nacht zum 24. Juli folgte Karamanlis diesem Ruf. Alsbald formte er eine sogenannte Regierung Nationaler Einheit, die aus einer eindrucksvollen Reihe angesehener Persönlichkeiten bestand und ein sehr breites politiches Spektrum repräsentierte. Ausgelassen wurde im Grunde nur die extreme Rechte im Sinne von Junta-Anhängern und die extreme Linke.

Es wäre grundfalsch, die Rolle des Volkes in diesem Wandel, der die Geburtsstunde der postdiktatorischen Freiheit darstellt, als nichtig anzusehen. überblickt man aber die Vielfalt materieller und moralischer, Inland und Ausland einbeziehender mitbewirkender Faktoren, so ist es unmöglich, den Wandel als Willensakt *eines* Souveräns, sei es eines souveränen Volkes zu betrachten.

Den formellen Übergang von der Diktatur zum neuen politischen Status leitet ein Dokument ein, dessen politische und rechtliche Verbindlichkeit nur aus dem Zusammenströmen mehrerer Elemente in der Ausnahmesituation von 1974 hervorgeht. Es ist ein sogenannter konstituierender Erlaß, also ein Erlaß des Präsidenten der Republik von verfassungsrechtlichem Rang. Es ist in der Form ein von der Diktatur oft verwendetes Instrument, um sich sogar über die von ihr selbst verkündete Verfassung hinwegzusetzen. Ein erstes relevantes Element des konstituierenden Erlasses des 1. August 1974 ist die Tatsache, daß er über der Unterschrift des letzten unter der Diktatur bestellten Präsidenten der Republik, General Ghizikes, erlassen wurde[5]. Die Vielfalt der bewirkenden und legitimierenden Faktoren des durch den Erlaß signalisierten Wandels werden dann auch in einer Präambel mit der Funktion einer Legitimationsformel mit beachtenswertem Feingefühl einbezogen. Ich komentiere den Text laufend. Er fängt an:

"Angesichts des am 23. Juli stattgefundenen Wandels. . ."[6]

Der Wandel wird nicht näher charakterisiert. Diese Zurückhaltung ist bereits ein Indiz seiner Vieldeutigkeit. Die Präambel fährt fort:

"und der durch diesen erfolgte Übertragung des Premierministeramtes an Konstantin Karamanlis. . ."[7]

Die ausdrückliche Verknüpfung des Wandels mit einer Person in einem konstitutiven Dokument von Verfasungsrang ist bemerkenswert und deutet auf die große Bedeutung, die den Qualitäten oder zumindest dem Image eines Mannes in diesem Zusammenhang zukam, der – auch im Ausland angesehen – das Vertrauen sowohl der Offiziere als auch der zivilen Öffentlichkeit genoß. Im weiteren heißt es:

"(Angesichts) des universalen und einstimmigen Beistandes des Griechischen Volkes;. . ."[8]

Die Realität dieses Beistandes und seine legitimierende Kraft braucht, was die große Mehrheit der Bevölkerung betrifft, nicht bezweifelt zu werden. Am 1. August ist er aber noch durch keinen formellen Konsultationsakt festgestellt worden.

Die innere und äußere Krise spielen in den nächsten Satz der Präambel hinein. Dieser lautet:

"Im Bewußtsein der schweren historischen Verantwortung, die die Regierung gegenüber der Nation auf sich genommen hat"[9]. Die Regierung ist die bereits der Öffentlichkeit bekannte Regierung Nationaler Einheit, die in ihrer personellen Zusammensetzung eine repräsentative Meinungsvielfalt darstellt, Kontinuität und Erneuerung verspricht und in vielerlei Hinsicht Vertrauen erweckt. Der nächste Satz ist ein rhetorisches Meisterstück:

"Inspiriert vom Patriotismus des Volkes und der hohen Moral und Macht der Verteidiger der nationalen Integrität und Sicherheit. . ."[10]

Die Macht der Armee bleibt unvergessen, zugleich aber nunmehr nach außen gerichtet, und der Patriotismus des Volkes scheint ihr darin zu folgen. Doch ist der Patriotismus auch jene Tugend, der der letzte Paragraph der Verfassung von 1952 die Erhaltung der Verfassung anvertraut – eine Bestimmung die auch als Widerstandsrecht ausgelegt worden ist[11].

Ich brauche die Formeln der Präambel nicht im Einzelnen weiter auszuspinnen. Zu den bereits erwähnten Legitimationselementen kommen der Ausblick auf die volle Wiederherstellung freien demokratischen Lebens und die entsprechende Selbsteinschränkung der Regierung als Provisorium *rem publicam restituendam* hinzu.

Der Konstituierende Erlaß des 1. August 1974 setzte die von der Diktatur abgeschaffene Verfassung von 1952 wieder ein, mit Ausnahme jedoch der sich auf die Monarchie beziehenden Artikel. Die Reaktivierung dieser Institution wurde vielmehr vom "freien Ausdruck des Willens des Griechischen Volkes"[12] abhängig gemacht. Neben anderen Übergangsbestimmungen ordnete der Erlaß einen Zusatz zu der bereits um die Monarchieparagraphen veränderten Verfassung von 1952 an. Die Sprache dieses Zusatzes unterstreicht die Absetzung des neuen politischen Status von der Diktatur, unter Verwendung des Begriffs von der Volkssouveränität:

"Kein Individuum oder Teil des Volkes darf die Ausübung der Volkssouveränität und ihrer Gewalten in irgendeiner Weise usurpieren"[13].

Der Zweite zur neuen Legalität überleitende, ebenfalls von General Ghiziks unterschriebene Konstituierende Erlaß vom 30. Oktober 1974 sieht vor,

1) daß innerhalb einer gegebenen Frist Parlamentswahlen ausgeschrieben werden,

2) daß die aus dem so gewählten Parlament hervorgehende Regierung innerhalb einer ebenfalls gesetzten Frist ein Referendum zur Frage der Monarchie abzuhalten habe,

3) daß das zu wählende Parlament befugt sein würde, alle Verfassungsbestimmungen inklusive der Grundlegenden zu revidieren, mit Ausnahme der durch das Referendum bis dahin entschiedenen Frage der monarchischen oder republikanischen Staatsform,

4) daß dem Parlament "zur Erleichterung seiner Revisionsarbeit" ein Verfassungsentwurf von der Regierung vorgelegt werden müßte,

5) daß es selbstverständlich der gewählten Kammer obliege, alle Bestimmungen des Entwurfs – bis auf die ausgeklammerte Staatsform – nach ihrem Gutdünken zu ändern, daß aber

6) im Falle daß das Parlament die Bearbeitung des Verfassungstextes innerhalb einer Frist von drei Monaten nicht abgeschlossen haben sollte, die Regierung befugt und angehalten sei, ihren Entwurf mit den bis dahin vom Parlament verabschiedeten Änderungen der Bestätigung durch ein Referendum auszusetzen[14].

Keine dieser, den Gang der Verfassungsentstehung maßgeblich vorzeichnenden Bestimmungen kann auch nur formal als Ausdruck des Volkswillens gedeutet werden, es sei denn als virtueller Auftrag demokratische Institutionen wiedereinzuführen.

Daran ändert nichts die Tatsache, daß diese Bestimmungen auf eine künftige direkte oder indirekte Konsultation der Wählerschaft vorausweisen.

Es ist unbestritten, daß die Volksmeinung zu den wesentlichen Faktoren der Verfassungsgestaltung und ihrer Legitimation gehört. Das tut sie bereits in diffusen Ausdrucksformen, die von Kaffee- und Salongesprächen bis zu monumentalen Manifestationen reichen, wie diejenige am Abend des 23. Juli 1974, an dem unzählige Menschen alle Straßen zwischen dem Zentrum von Athen und dem Flughafen stundenlang füllten, um Karamanlis, der die Restauration der Freiheiten symbolisierte, zu empfangen. Das tut sie auch in der institutionell artikulierten Form von Referendum und Wahl. Aber die Volksmeinung ist eben das: reagierende Meinung, nicht agierender Wille. Das Urteil des Volkes bleibt – bis auf spontane Demonstrationen, deren repräsentative Qualität selten eindeutig und deren Gehalt notwendig undifferenziert ist – latent, bis es gefragt

wird. Das Volk als solches initiiert nichts. Es gibt seine Zustimmung oder verweigert sie. Es zieht diesen aufgestellten Kandidaten jenem vor. So kann und soll man im Volk eine Kontrollinstanz erblicken, an der nicht vorbeiregiert werden darf. Muß aber nicht ein stärkeres Moment originären Wollens dabei sein, um von einem Souverän sprechen zu können[15]?

"Es ist ein radikaler Unterschied" schreibt Siegfried Landshut, "für den organisatorischen Aufbau, die Verfassung einer politischen Gemeinschaft, einmal: ob keine Entscheidung fallen darf, die nicht in Übereinstimmung mit der öffentlichen Meinung steht. In diesem Fall ist es nicht Sache der öffentlichen Meinung, einen Willen im Hinblick auf eine bestimmte Entscheidung zu produzieren. Dies tun andere – aber die öffentliche Meinung reagiert, und ihre negative Reaktion muß auf die Dauer fähig sein, die getroffene Entscheidung zu annullieren. Oder: ob der entscheidende Wille selbst, der Entwurf des zu realisierenden ein Produkt des Volkes selbst sein soll. In diesem Fall bedeuten "unpopuläre Maßnahmen" . . ., die sich die Politiker sogar zum Verdienst anrechnen, nichts als: Ungehorsam gegen den Souverän"[16].

Der breite Konsens des griechischen Verfassugsrevidierenden Parlaments von 1975 bezog sich in seiner Betonung der Volkssouveränität zweifellos vornehmlich auf die Gewährleistung und Absicherung der *Kontroll*befugnisse des Volkes. Die Euphorie der Demokratischen Restauration und der Wunsch, sich von der gestürzten Diktatur abzusetzen, begünstigte jedoch die "klassischen" voluntaristischen Konnotationen des Souveränitätsbegriffs. Ich möchte hier nicht über künftig mögliche Auslegungen der Volkssouveränitätsformel spekulieren, sondern nur auf Tendenzen hinweisen, die bereits in der Verfassungsdebatte aufgetreten waren. Die lokale Selbstverwaltung wurde z.B. nicht nur im Interesse der politischen Partizipation, sondern, als wäre es dasselbe, als "Originäre Manifestation" der Volkssouveränität reklamiert[17]. Die konstitutionelle Absicherung der Wahlbeteiligung von Staatsbürgern, die sich im Ausland oder auf hoher See befinden, wird nicht etwa im Hinblick auf die angemesene Vertretung ihrer besonderen Interessen, sondern im Namen der Volkssouveränität verlangt[18]. Hier wird jene magische Kraft evoziert, die die Vielfalt partikularer Meinungen in einen authentischen und darum wahren und gerechten Willen transfiguriert, wenn nur alle Meinungsträger ungehindert zusammenkommen. Und obwohl das Wirken eines solchen Wunders im Grunde keine Abstufungen zuläßt, so werden Approximationen dennoch angestrebt. So wurde – vornehmlich von der Opposition – im Namen der im Artikel 1 zur Grundlage der Verfassungsordnung erklärten Volkssouveränität, für das proportionale Wahlrecht, für eine schwache Exekutive usw. argumentiert[19]. Ich kann auf diese Probleme, die einer eingehenden Analyse bedürfen, hier nur hinweisen.

Es ist häufig zu Recht bemerkt worden, daß die voluntaristisch-absolute Ausprägung des Souveränitätsbegriffs seiner Geschichte als

Kampfbegriff im Dienst von Monarchen, Parlamenten oder Revolutionen entstammt. Zugleich ist für die "klassische" Souveränitätsauffassung die Frage nach dem Träger von zentraler Bedeutung[20].

Es ist beachtenswert, daß im Bereich, wo die Absetzung gegen die Diktatur eine Einschränkung der Souveränität erforderte, die Frage des Trägers indifferent erschien. So ist weiterhin im Artikel 28 der jüngsten griechischen Verfassung, wo es darum geht, die Verbindlichkeit übernationaler Normen, wie Menschrechtsdeklarationen für das interne Recht anzuerkennen, von der Souveränität der Nation die Rede[21].

Im bestimmten Kampfton der Distanzierung von der Diktatur wurde 1974 die Volkssouveränität zum Schlagwort nicht nur in den offiziellen Verlautbarungen sondern auch in der Presse. Die Volkssouveränität sei das, um dessen Restitution man sich freue und um dessen Schutz man besorgt sei, auch wenn die Vorstellungen über die institutionelle Ausfüllung dieser Formel sehr weit auseinanderlagen[22].

Die Anziehungskraft der Formel von der Volkssouveränität in der Situation von 1974 liegt nicht zuletzt in ihrer Fähigkeit, in der streitbaren Abgrenzung gegen die Diktatur auf der Sloganebene vereinheitlichend zu wirken.

Vor diesem sprachlich zugunsten der Volkssouveränität bereits kolorierten Hintergrund entfachte sich die Debatte um die Formulierung von Artikel 1 der Verfassung. Die Wahlen des 17. November 1974 hatten der "Neuen Demokratie", der Partei von Karamanlis 54,37% der Stimmen und 220 von 300 Parlamentssitzen eingebracht. Der Zentrumsunion fielen 60 Sitze zu; der Panhellenischen Sozialistischen Bewegung von Papandreou 12. Zwei kommunistische Parteien und andere linke Splittergruppen, die sich im Wahlkampf als "Vereinigte Linke" präsentiert hatten, teilten sich die übrigen 8 Sitze. Formationen, die extrem monarchistische oder diktaturfreundliche Elemente vertraten, erhielten knapp über 1% der Stimmen und keinen Parlamentssitz. Die Karamanlisregierung besaß also mehr als die zur Verfassungsrevision benötigte Zweidrittelmehrheit. Die Abweichung von ihrem Verfassungsentwurf in der Formulierung des ersten Artikels ist angesichts dieser Tatsache besonders bemerkenswert. Fast wichtiger jedoch für unsere Fragestellung als die politische Zusammensetzung des Parlaments erscheint die Dominanz legalistischen Geistes, die sich nicht nur an der Zahl von Anwälten unter den Abgeordneten, sondern auch und vor allem in der formal staatsrechtlichen Schulung der Protagonisten der Verfassungsdebatte manifestiert. Die Stellung von Verfassungsfragen als wären sie primär, wenn nicht ausschließlich Rechtsfragen, wirft Probleme auf, die der Kritik von einem politikwissenschaftlichen Standpunkt in besonderem Maße bedürfen. Die Suche nach einem einzigen, sei es auch kollektiven Souveränitätsträger geht auf die Ansicht zurück, daß die im Verfassungsdokument verbriefte Gestaltung von Gewalten, Absicherung von Rechten, Festlegung von Prozeduren usw. nichts

als eine Art höheren formalen Gesetzes darstelle und daß man deshalb denjenigen finden müsse, der dieses Gesetz spricht. Doch die Verfassung ist keinewegs das Gebot eines per definitionem dazu befähigten Gebieters. Verlangt man vom Souveränitätsbegriff mehr als rechtstechnisches Instrument der Abgrenzung von Jurisdiktionen zu sein und wird damit die Frage nach dem Träger für die inhaltliche Bedeutung des Begriffs maßgeblich, so wird in der Suche nach dem Inhaber konstituierender Dezisionsmacht die Vielzahl der materiellen Machtinstanzen und normativen Kompetenzen verkannt, die in aller Regel bei der Verfassungsentstehung zusammenwirken. Wo der Primat einer rechtlichen Verfassung, wie 1975 in Griechenland, nicht glaubhaft postuliert werden kann, und dennoch von einer rechtsformalen Betrachtungsweise ausgegangen wird, ist die Idee einer *persona ficta* als immerwährenden Trägers des *pouvoir constituant* außerordentlich verlockend.

Die Bezeichnung der Verfassungsgebung von 1975 als Verfassungsrevision täuscht eine Rechtskontinuität vor, die aber in Wirklichkeit nicht vorhanden war.

Der Konstituierende Erlaß vom 1. August kann weder formal noch inhaltlich als Akt der "Wiederbelebung" einer während der Diktatur latenten, aber nun wieder im vollen Sinne gültigen vordiktatorischen Verfassung ausgelegt werden, sondern lediglich als politisch wirksame Notlösung in einem Rechtsvakuum. Man kann unmöglich aus der Verfassung von 1952 Kompetenzen von Präsident Ghisikis ableiten. Die substanzielle Änderung des Verfassungstextes per Dekret entspricht wohl auch kaum den in der Verfassung vorgesehenen Revisionsbestimmungen, zumal diese die Änderung "Grundlegender Artikel" – zu denen man wohl diejenigen über die Staatsform rechnen muß – überhaupt untersagte[23]. Um eine weitere Komplikation hinzuzufügen, war die Verfassung von 1952, die die Revision einer früheren darstellt, ebenfalls nicht nach den in ihrer Vorgängerin vorgesehenen Prozedur zustandegekommen. Und so kann man die Brüche bis zum Anfang der griechischen Verfassungsgeschichte zurückverfolgen, einem Anfang der selbst in der Emanzipation von der Osmanischen Herrschaft revolutionär gesetzt war. Die rechtliche Fiktion einer Verfassungsentstehung *ex nihilo* verlangt die rechtliche Fiktion eines souveränen, archimedischen Hebelpunktes. Eine solche Konstruktion wird problematisch, wenn man einen solchen Punkt *in* der Welt der politischen Konkretion sucht.

Konflikte dieser Art sind nicht nur im Bezug auf die Hervorhebung der Souveränitätsformel, sondern auch in der Argumentation um den angemessenen Träger aufgetreten.

Gegen die traditionelle Formulierung von der Nation als Quelle aller Gewalten haben alle Oppositionsparteien, aber auch viele Abgeordnete der Regierungsmehrheit polemisiert und Änderungsanträge eingereicht. Eine Seite der Argumentation war historisch:

Der Nationsbegriff sei dadurch belastet, daß im Namen der Nation Rechte und Freiheiten des Volkes eingeschränkt worden wären, wie an der soeben gestürzten Militärherrschaft exemplifiziert werden konnte, die sich Nationale Revolution zu nennen pflegte[24].

Zur jüngsten Diktaturerfahrung wurde je nach politischer Einstellung unterschiedlich stark die gesamte Nachkriegszeit gezählt, während der, unter dem Schatten eines vierjährigen Bürgerkrieges, die "Nation" zum Symbol des Antikommunismus wurde[25].

Auf der anderen Seite wurde argumentiert, daß alle Worte dem Verschleiß unterliegen. Die Worte Demokratie und Sensibilität z.B. nicht weniger als das Wort Nation und daß die historische Belastung des Volksbegriffes durch die Errichtung von sogenannten "Volksgerichten" im Bürgerkrieg nicht geringer sei als der Schaden, den Deportationen unter dem Gesichtspunkt der "Nationalen Gesinnung" dem Nationsbegriff zugefügt habe[26]. Der wirkliche Gewinn dieser Debatte liegt jedoch darin, daß es gelungen ist, die Diskussion aus diesem Niveau herauszuheben und eine sich symbolisch an diesem Begriffstreit angehängte politische Polarisierung zu verhindern, die die Schlagworte des Bürgerkrieges als solche wiederaufgegriffen hätte. Die Klarstellung in der Souveränitätsformel, die durch den hinzugefügten Absatz:

"Die Volkssouveränität ist das Fundament der Verfassungsordnung" erfolgte, geht auf die Initiative des Vorsitzenden der Verfassungskommission, der in Kürze Präsident der Republik werden sollte, zurück, eines unzweideutigen Repräsentanten der Seite, die den Bürgerkrieg gewonnen hat, der Seite der Nation[27]. Die endgültig in die Verfassung aufgenommene Souveränitätsformel, die in Kontinuität zur Sprache der Konstituierenden Erlasse und der öffentlichen Diskussion steht, ist kein Kompromiß – der angesichts der Regierungsmehrheit unnötig gewesen wäre – sondern das Zeichen einer negativ durch die Diktaturerfahrung angebahnten Versöhnung.

Die Fürsprecher des Volksbegriffs argumentierten, daß der Volksbegriff nicht nur aus historisch kontingenten, sondern auch aus prinzipiellen Gründen dem Nationsbegriff vorzuziehen sei[28]. Um die Widerstandskraft des Volksbegriffs gegen ideologische Verklärung aufzuzeigen, folgten sie zwei in den Reden verflochtenen in der Sache aber widersprüchlichen Argumentationslinien. Im Gegensatz zur diffusen ideellen Realität "Nation" sei das Volk – sagten sie – eine örtlich und zeitlich bestimmte, faßbare, d.h. soziologisch auffindbare Realität. Zugleich – und dies ist die andere, vom Juristengeist geprägte Argumentation, sei das Volk mit staatsrechtlicher Genauigkeit zu erfassen und habe, im Gegensatz zur Nation – eine institutionell konkretisierbare Willensäußerung.

Unabhängig von der auch unter Juristen strittigen Frage, ob die mit originärer verfassungsgebender Kompetenz versehene Volkssouveränität rechtlich faßbar ist, ist sicher, daß das "Volk", das hier als ihr Träger

auftritt, ein anderes ist als das konstituierte Wahlvolk. Dem konstituierten Wahlvolk kommen die Qualitäten in der Tat zu, die die juristisch argumentierenden Fürsprecher des Volksbegriffs für ihren Souverän in Anspruch nehmen: die Staatsorgansqualität, die juristisch saubere Bestimmung und die notwendig nachkonstitutionelle, institutionalisierte Ausdrucksmöglichkeit. Die Identität von Schöpfer und geschaffenem Organ scheint mir schon deshalb unmöglich, weil das konstituierte Wahlvolk von einer Reihe konstitutioneller und abgeleiteter Faktoren wie Bürgerrecht, Mündigkeitsalter, Wahl- und Registrationsrecht usw. abhängig ist, der vorkonstitutionelle Souverän ist dagegen völlig unbestimmt.

Die Suche nach empirischen Willensträgern – der hypostasierungsverhindernden Konkretion zuliebe – führt zu einer Vielzahl konkreter Menschen, die erst als juristische Person zusammengefaßt, das heißt, von ihrer Individualität abstrahiert gedacht werden müssen, um der ebenfalls gesuchten staatsrechtlichen Präzision zu genügen.

Zur juristischen Person gehört aber auch eine Kontinuität in der Zeit, die einem Souverän abgeht, dem man jegliche Transzendenz aus Furcht vor Hypostasen abspricht. Hier aber gelang es, in der Debatte, das Volk als konkret augenblicklich präsenten Willens- und damit Souveränitätsträger hinzustellen, als jeweilige Inkarnation der Nation, die wiederum als nichts außer das Volk in seiner historischen Kontinuität dargestellt wurde[29].

Schwieriger als das Problem der zeitlichen war das der örtlichen Transzendenz der Nation. Das außerhalb des Territoriums befindliche Griechentum kann im Gegensatz zu vergangenen und künftigen Generationen nicht als momentane Manifestation eines Kontinuums, sondern nur als *Teil* der Nation auftreten. Geht man aber weiterhin von der Vorstellung einer im Willen konkreter Träger konstituierten Souveränität aus, so ist es konsequent, die griechische Diaspora als Quelle der Gewalten der griechischen Republik auszuschließen. Dies aber bedeutet die Ausschließung von der Souveränitätsformel – dem einzigen in der Griechischen Verfassung auffindbaren *fons honoris* – einer für das Identitätsbewußtsein der Republik – eben als griechischen Nationalstaates – determinierenden Einheit[30]. Daher stammt das in der Debatte artikulierte Verlangen nach der Beibehaltung der Nation als "breiteren" und "umfassenderen" Souveränitätsträger. Das Problem ist aber nicht, daß der Souvernänitätsträger zu eng gefaßt ist, sondern, daß die essentiell voluntative Souvernänität ungeeignet ist, eine nicht voluntativ konstituierte Einheit, die *außerdem* das Staatsvolk zeitlich und örtlich überflügelt, zu symbolisieren.

Die Formulierung des endgültigen Verfassungstextes scheint mir dieses Problem trotzdem, soweit es die gegebenen Kategorien zulassen, nicht unbefriedigend gelöst zu haben: Die dem Volk entspringenden Gewalten werden zugunsten des Volkes und der Nation ausgeübt[31]. Nicht als Quelle

der Gewalten, sondern als erstrangiger Gegenstand der Fürsorge des Souveräns kann die "Nation" Dimensionen griechischen Lebens repräsentieren, die die unmittelbare Gegenwart und die Grenzen der Republik übersteigen. Eine solche Formel bedeutet zugleich eine förmliche Absage an den Irredentismus, dem nationalistischen Streben Staatsvolk und Nation koextensiv zu machen, der die Politik des griechischen Staates seit seiner Gründung 1830 begleitet und zeitweilig dominiert hat. Insoweit scheint die Souveränitätsformel der Verfassung von 1975 im Rahmen ihrer an sich problematischen Symbolsprache eine historische Hypothek griechischer Existenz, auf der für Bewußtseinsrealitäten signifikanten symbolischen Ebene endgültig getilgt zu haben.

## Anmerkungen

[1] Die Protokolle und Dokumente der Verfassungsdebatte sind in drei Bänden versammelt worden:
Parlament der Hellenen - Fünfte, mit der Verfassungsrevision beauftragte Nationalversammlung
a) Protokolle der Unterkommissionen des parlamentarischen Verfassungsausschusses 1975, Athen, Juli 1975 (zitiert: Unterkommissionen).
b) Protokolle des Plenums des Verfassungsausschusses, 1975 (zitiert: Plenum).
c) Protokolle der Parlamentssitzungen zur Verfassungsdebatte 1975, Athen, 1975 (zitiert: Parlament).

[2] Parlament, 12, von der Regierung eingebrachter Verfassungsentwurf, S. 1145, Art. I,2.

[3] Parlament, 15, Entgültiger Verfassungstext, S. 1219, Art. I, 2,3.

[4] Siehe jedoch in Bezug auf die *praktische* Bedeutung der allgemeinen Verfassungsbestimmungen jenseits ihrer symbolischen Kraft, D. Tsatsos: "Sind die allgemeinen Verfassungsbestimmungen bloß eine Selbstdarstellung des Staates, der sich nach außen als demokratisch, parlamentarisch, als Rechtsstaat gibt? Nein. . . sie haben auch einen konkreten praktischen und rechtlichen Wert. . . denn der Gesetzgeber muß häufig auf allgemeine Maximen der Rechtsordnung zurückgreifen, wenn er nicht in der Lage ist, besondere Fälle auf der Grundlage besonderer Bestimmungen zu entscheiden." Plenum, 11/3/75 p. 255.

[5] Parlament, 2. Konstituierender Erlaß vom 1.8.1974, I,2.

[6] l.c.

[7] l.c.

[8] l.c.

[9] l.c.

[10] l.c.

[11] Zur Ambiguität des Art. 114 der Verfassung von 1952 siehe: Chr. Sgouritsas, Verfassungsrecht (auf Griechisch) o.o., o.J. vol. I, s. 70 ff. der auch auf Art 169 des Strafgesetzbuches verweist, der die Verweigerung eines rechtswidrigen Dienstes vorsieht, sowie auf den Beamtenkodex Art. 45,3 bezüglich der Nichtausführung eines rechtswidrigen Befehls. Siehe auch die Rede von G. Mangakis, Unterkommission, 29.1.75, s. 419.

[12] Art. I,2 Konstituierender Erlaß vom 1.8.74, Parlament s. 1. Von der Wiederinkraftsetzung werden ausgeschlossen Art. 433; 45-53 der Verfassung von 1952.

[13] Art. 6 1/8/74 l.c.

[14] Konstituierender Erlaß 3-4/10/74 o.c. s. 3,4.

[15] Siehe, u.a. Dolf Sternberger, *Nicht alle Staatsgewalt geht vom Volke aus*, Sutttgart 1971.

[16] Siegfried Landshut, "Volkssouveränität und öffentliche Meinung", in *Gegenwartsprobleme des internationalen Rechts und der Rechtsphilosophie. Festschrift für Rudolf Laun*, Hamburg 1952, s. 579 ff. nunmehr in: Hans Kurz Hrsg., *Volkssouveränität und Staatssouveränität*, Darmstadt 1970 (Wege der Forschung Bd. XXVIII) s. 306.

[17] Siehe Änderungsvorschlag zum Art. 102, des von der Regierung vorgelegten Verfassungsentwurfes 1975 von E. Eliou, Parlament, 14/5/75, s. 823: "Die lokale Selbstverwaltung ist ursprünglicher Ausdruck der Volkssouveränität. . .". Siehe auch die Rede von A. Papandreou, Parliament, 14/5/75, s. 823. und Th. Manavis, o.c. S. 835, 839-41.

[18] Siehe: D. Tsatsos, Plenum, 25/2/75, s. 102. Th. Manavis nimmt "Volk" im engen Sinne von "Wahlvolk" und schließt konsequenterweise diejenigen aus, die dieser juristischen Körperschaft als solcher nicht angehören. o.c. S. 105. Das Problem stellt sich durch die lexikographische, aber auch politische Vieldeutigkeit des Begriffs "Repräsentation".

[19] Es geht im Grunde um das im französischen Sinne, im Sinne von Alain, "Radikale" Argument, demnach der Bürger sich gegen die Gewalten verteidigt, indem er ihnen die öffentliche Meinung "Opinion" entgegensetzt, die, um exakt ausgedrückt und darum "wahr" zu sein, des Proporzes bedarf. Es ist mehr ein anfechtendes-herausforderndes Demokratiemodell, mehr denn ein partizipatorisches. Es ist bemerkenswert, daß in der Debatte Argumente zugunsten von Wahlrechtsvorkehrungen angeführt sind, die die größeren Parteien und Formationen favorisieren, nur im Sinne der Regierungsstabilität und Effizienz angeführt werden und daß nirgends der Gedanke aufkommt, daß solche Vorkehrungen zur Integration und Programmbestimmung bereits auf der Ebene der Wähler begünstigt werden könnte, statt sich dem Kuhhandel und den Austauschmehrheiten in der Kammer anzuvertrauen. Die klassischen Werke dazu sind natürlich: Herman Finer, *The Case against Proportional Representation*, Fabian Tract Nr. 211, London 1935 und in seinem Gefolge Ferdinand A. Hermens, *Democracy or Anarchy? A Study of Proportional Representation*, Notre Dame 1941.
Andererseits verdienen die durch das Mehrheitswahlrecht oder die grossen Parteiformationen begünstigende Wahlsysteme aufgeworfenen Probleme, vor allem in Situationen ohne umfassenden, einrahmenden Konsens, einer eingehenden Studie. Die ausgeprägte Polarisierung des griechischen politischen Lebens ist nur ein Beispiel unter vielen. Die Probleme stellen sich bereits im Heimatland des parlamentarischen Systems selbst, Großbritannien. Siehe: S.F. Finer, ed. *Adversary Politics and Elctoral Reform*, London 1975. Wie dem auch sei, scheint demnach der Proporz kein notwendiges Postulat der Volkssouveränität zu sein. Es muß auch zugegeben werden, daß die meisten Redner nur Garantien gegen einen allzu großen "Bias" verlangten. Siehe z.B.: D. Tsatsos, Parlament, 3/4/75, s. 117; A. Papandreou o.c. s. 123; K. Kappos, o.c. s. 142.

[20] Siehe: Peter Häberle, "Zur gegenwärtigen Diskussion um das Problem der Souveränität" in: Archiv des öffentlichen Rechts, Bd. 92, 1, 67 ss. 259-287.
Jedoch, im Gegensatz zur Vielfalt von Faktoren, die – nach Häberle – zusammen zur normativen Begründung der Republik beitragen, finden wir in der griechischen Verfassungsdebatte die Suche nach einer einheitlichen "Kompetenz-Kompetenz" eg.: Th. Manavis, Plenum, 11/3/75 s. 265; wiederaufgegriffen: Parlament, 31/3/75, s. 78; aber auch in anderen Formulierungen.

[21] Art. 28,3, Verfassung 1975, Parlament s. 1226.

[22] Es ist unmöglich, die Pressestimmen der Zeit hier erschöpflich zusammenzutragen. Einige Hinweise, auf direkte Behandlungen der "Nation oder Volk" als Souveränitätsträger oder Quelle aller Gewalten müssen genügen: *To Bema*: 5/1/75, G.A. Koumandos; 11/1/75 A.P. Kanellopoulos; 12/1/75 G. Tenekides; 12/1/75 M. Plorites; 9/2/75 "Studiengruppe für eine demokratische Verfassung", 1/10/75 D. Tsatsos: *Kathemerine*: 4/3/75; A. Floros *Vradini*: 1/4/75 J. Theodorakopoulos.

[23] Art. 108, Verfassung 1952.

[24] A. Sehiotis, Unterkommission 24/1/75, s. 65; 30/1/75 s. 118. Er fügt mit einem vielleicht übertriebenen Optimismus hinzu, daß Diktatoren den Volksbegriff schwerlich ausnützen könnten. Siehe die skeptischen Überlegungen diesbezüglich von I. Varvitsiotis, Unterkommission, 20/1/75, s. 114, der nicht nur vom demagogischen Gebrauch des Volksbegriffs durch Mussolini, Hitler, Stalin etc. spricht, sondern auch auf die Herleitung der Gewalten vom Volk gemäß Art. 2., der von der griechischen Diktatur proklamierten Verfassung von 1968, hinweist. Siehe auch A. Michas, Plenum, 25/2/75. s. 106 und M. Ploritis in *To Bema*, 12/1/75, s. 12.

[25] E. Eliou, Unterkommission, 24/1/74; 30/1/75. Zu den die Begriffe belastenden historischen Hypotheken siehe auch: D. Nianias, Plenum, 25/2/75, s. 106, D. Tsatsos, o.c., 11/3/75, s. 255.

[26] K. Tsatsos, Plenum 25/2/75, s. 107. Auch P. Kanellopoulos, o.c. s. 109.

[27] K. Tsatsos, Plenum, 11/3/75, s. 264, 267. Und mit Nachdruck: Parlament 31/3/75, s. 68.

[28] Zum "Volk" als präzisen juristischen Begriff, der einer konkreten Realität entspräche und damit keine Mystifikationen zuläßt siehe: G. Koumandos, *To Bema* 5/1/75; A.P. Kanellopoulos, *To Bema* 11/1/75; Plenum 25/2/75 ss. 105, 106, 108; N. Alavanos Unterkommission 24/1/75, s. 69; E. Elion o.c. s. 59; K. Paparigopoulos, Plenum 25/1/75, s. 105; N. Papapolitis o.c. s. 106; K. Tsatsos o.c. s. 107; G.B. Mangakis o.c. s. 107; A. Sehiotis o.c. 107/108.
Dagegen habe ich lange über ein Argument zugunsten der Beibehaltung der "Nation" als Quelle der Gewalten nachgedacht, ohne, muß ich zugeben, seinen Sinn ergreifen zu können: Nämlich daß der Nationsbegriff "erbaulich" sei: K. Paparigopoulos, Plenum 25/2/75, s. 105.

[29] K. Tsatsos, Unterkommission, 24/1/75, s. 67: "denn das Volk des jetzigen Augenblicks sorgt für die entfernteste Zukunft. . ."

[30] Andererseits können "nationale" Minderheiten wohl in ein souveränes Volk, nicht aber in eine souveräne Nation aufgehen wenn "Nation" romantische Konnotationen mitschwingen – Konnotationen, die es im zeitgenössischen friedlichen Bewußtsein nicht genug abgelegt hat. Die Revidierende Nationalversammlung hat sich natürlich explicit geweigert, griechische Staatsangehörige von der Präsidentschaft der Republik – und a *fortiori* von geringeren Chargen – auszuschließen auf Grund ihres ethnischen, kulturellen oder religiösen Ursprungs oder Hintergrunds, auch wenn es derartige Vorschläge gegeben hat, die einen Skandal darstellen würden, wenn sie nicht lächerlich gewesen wären.
Zur "virtuellen" Repräsentation der Auslandsgriechen durch das griechische Parlament bzw. durch die griechische Regierung siehe:
Th. Manavis, Plenum, 25/2/75, s. 105;
A. Minas Unterkommission 22/1/75, s. 328.
Dagegen, in Bezug auf die überwindung des Irredentismus:
I. Varvitsiotis, Unterkommission, 30/1/75, s. 115.
Zur Diskussion über die Diaspora:
D. Chloros, Unterkommission, 24/1/75, s. 60;
N. Papapolitis, o.c. s. 65;
A. Katsaounis, 21/1/75, s. 357;
A. Kalatzakos s. 349;
K. Tsatsos 24/1/75, s. 67;
A. Sehiotis s. 65;
I. Varvitsiotis s. 63; 30/1/75, s. 113;
N. Papapolitis s. 64;
G.B. Mangakis, Plenum, 25/2/75, s. 107.

[31] Art. I, 2,3, Verfassung 1975.

# Geheimnis
# Ein interaktionistisches Paradigma

BIRGITTA NEDELMANN
Freiburg/Florenz

"Das Geheimnis, das durch negative oder positive Mittel getragene Verbergen von Wirklichkeiten, ist eine der größten Errungenschaften der Menschheit" (Simmel 1968, 272)

Ebenso wie der Anspruch auf Umverteilung materieller Güter kann auch der auf Umverteilung von Wissen zum Gegenstand sozialer und politischer Auseinandersetzungen werden. Insbesondere solche Fälle bergen ein hohes Konfliktpotential, in denen Wissen monopolisiert und anderen daran Interessierten bewußt und absichtlich vorenthalten wird. Es sind dies Fälle, in denen durch Geheimhaltung Macht geschöpft und ergriffen wird. Diese Akte ziehen meistens dann heftige Turbulenzen nach sich, wenn sie ganz offensichtlich Normen zur Veröffentlichung von Wissen verletzen und außerdem von solchen Akteuren begangen werden, die ohnehin über Machtressourcen verfügen. Dies trifft vielfach für Positionsinhaber im Politikbereich demokratisch verfaßter Gesellschaften zu. Die Möglichkeit zur Kontrolle der gewählten Vertreter durch Teilhabe an Wissen ist zentraler Bestandteil dieser Systeme. Diese Kontrollmöglichkeit wird ganz offensichtlich dann vereitelt, wenn Bürgern der Zugang zu für sie relevantes Wissen versperrt wird, aber auch dann, wenn auf sie zu viel für sie irrelevantes Wissen abgeladen wird.

Im Mittelpunkt derzeitiger Zeitdiagnose stehen zwei Vermutungen über die Entwicklungstendenzen von Geheimnis: Zum einen wird der Verdacht geäußert, Geheimnis nehme im Politikbereich immer mehr zu; das Öffentliche werde immer privater. Zum anderen wird der Verdacht einer gleichzeitig stattfindenden, gegenläufigen Tendenz geäußert, nach der die Geheimhaltung privater (oder intimer) Angelegenheiten immer mehr abnehme; das Private werde immer öffentlicher (so etwa kürzlich Guggenberger 1985). Sollte – um den von Richard Sennett geprägten

Ausdruck zu übernehmen – die "Tyrannei der Intimität" die "Tyrannei der Öffentlichkeit" ersetzt haben? (Siehe hierzu Sichtermann 1985).

Georg Simmel hat in seiner im Jahre 1908 publizierten "Soziologie" eine sog. Evolutionsformel aufgestellt, die der heutigen Zeitdiagnose widerspricht. Danach wird "das Öffentliche immer öffentlicher, das Private immer privater" (Simmel 1968, 277). Haben die jüngsten Entwicklungen Simmels Evolutionsformel überholt? Oder beruht die heutige Kritik auf einer Überschätzung der durch Massenmedien hochgespielten Skandalfälle und Affären? Diese Fragen lassen sich in der so gestellten Form nicht beantworten. Sie können auch nicht Diskussionsgegenstand eines Beitrages sein, der seinem Charakter nach nicht kulturkritisch sein soll. Er beschränkt sich auf den Versuch, eine wichtige Vorfrage zu klären: *Was ist eigentlich genauer unter Geheimnis zu verstehen*? Welche sozialen Beziehungen werden inszeniert oder abgeschnürt, wenn Wissen geheim gehalten wird? Die Klärung dieser Fragen erlaubt es, auf Widersprüche, Spannungen und Prozesse hinzuweisen, die durch Geheimhaltung als Interaktionsform als solche ausgelöst werden. Viele aktuelle politische Skandal-Fälle (von der Fälschung der Hitler-Tagebücher, über den Flick-Skandal und die Parteispendenaffäre, bis zum Weinskandal) ließen sich eher verstehen, wenn man Erkenntnisse über Geheimnis als Interaktionsform von Anfang an berücksichtigen würde.

Der Rekurs auf kontingente politische "Umstände" und "Bedingungen" reicht oft bloß aus, um die von Fall zu Fall inhaltlich wechselnden, aber funktional austauschbaren politischen Motive für Geheimhaltungsabsichten zu erfassen, nicht aber, um Struktur und Dynamik geheimer Konstellationen selbst zu begreifen.

Ein solcher *Rekonstruktionsversuch von Geheimnis als Interaktionsform* soll hier unternommen werden und zwar auf der Grundlage des fünften Kapitels aus Simmels "Soziologie":"Das Geheimnis und die geheime Gesellschaft".

Georg Simmel ist vorgehalten worden, Geheimnis nur unzulänglich definiert zu haben (so etwa von Sievers 1974). Diese – wie immer berechtigte – Kritik dient oft bloß als Vorwand, um sich der Mühe einer konstruktiven Erarbeitung der von Simmel hauptsächlich verwendeten analytischen Elemente von Geheimnis zu entziehen. Scheut man diese Mühe jedoch nicht und läßt sich von Simmel zu Ergänzungen und Weiterführungen seiner Gedanken anregen, so läßt sich ein sog. interaktionistisches Paradigma von Geheimnis erstellen.

Simmel analysiert Geheimnis unter der Perspektive der Wechselwirkung. Dies mag zunächst trivial erscheinen, hat aber Konsequenzen für die Definition und Auswahl der empirischen Fälle von Geheimnis. Als Soziologe interessiert Simmel Geheimnis hauptsächlich insofern, als es soziale Beziehungen stiftet. Er räumt ein, daß ". . .man mit Recht sagt, ein Geheimnis, um das Zwei wissen, sei keines mehr" (Simmel 1968, 283).

Die soziale Wirkung von Fällen, in denen ein einzelner Wissen für sich allein behält, ist "Isolierung, Gegensatz, egoistische Individualisation" (Simmel 1968, 282); sie sind unter der Perspektive der Wechselwirkung soziologische Grenzfälle. In dem Moment, in dem ein Akteur Wissen *bewußt* und *gewollt* angebbaren anderen Akteuren vorenthält *und* es mit angebbar anderen Akteuren teilt, – erst in diesem Moment entsteht eine Dreierkonstellation, die jene Spannung, Emotionalität und Dynamik enthält, die die Attraktivität und Ambivalenz von Geheimnissen ausmachen. Welches sind nun im einzelnen die analytischen Elemente des interaktionistischen Paradigma?

## I. Anzahl von Akteuren und Arten von Handlungen im Geheimnis

Folgt man Simmel und betrachtet Geheimnis unter der Perspektive der Wechselwirkung, so stellt sich zunächst die Frage, *wie viele* Akteure typischerweise am Geheimnis partizipieren und welche *Arten von Handlungen* sie dabei entfalten. Die Beantwortung dieser Fragen führt zum ersten Definitionsmerkmal:

Geheimnis stiftet typischerweise Triaden: Zwei Akteure, (A) und (B), *teilen* Wissen (I.1) und machen es dadurch zu einem exklusiven Gut, daß sie es vor Dritten, (C) *verbergen* (I.2). Der "stark betonte Ausschluß aller Draußenstehenden" verleiht den Geheinisträgern "ein entsprechend stark betontes Eigentumsgefühl" (Simmel 1968, 273). Teilen von Wissen assoziiert (A) und (B) noch in einer anderen Hinsicht: Sie verständigen sich darüber, welche Wissensbestandteile anderen vorenthalten werden sollen und elaborieren u.U. Symbole und Codes zur Binnenkommunikation. Die Beherrschung einer Geheimsprache hat mehr als nur instrumentelle Funktionen: (A) und (B) qualifizieren sich damit nicht nur als Träger exklusiven Wissens, sondern auch als Träger einer exklusiven Kultur.

Während die Teilhabe von Wissen assoziiert, hat das Verbergen von Wissen die Wirkung der Dissoziation. Geheimnisträger müssen eine Vorentscheidung darüber treffen, welche anderen Akteure von der Teilhabe an dem exklusiven Gut ausgeschlossen werden sollen. Durch Verbergen wird eine Schranke zwischen Wissenden und Nicht-Einzuweihenden konstruiert. Dadurch bricht aber die Wechselwirkung nicht ab; vielmehr werden in Orientierung an den ausgeschlossenen Dritten vielfältige Techniken des Verbergens erfunden, die ethisch mehr oder minder bedenklich, in manchen Fällen sogar paradox sein können. Plagen den Geheimnisträger Schuldgefühle, weil er vor dem Dritten etwas verbirgt, dann wird er "dadurch zu Rücksicht, Zartheiten, geheimem Wiedergutmachenwollen bewegt, zu Nachgiebigkeiten und Selbstlosigkeiten, die ihm bei völlig

gutem Gewissen ganz fernlägen" (Simmel, 1968, 272, Anm.1). Verbergungstechniken betreffen zum einen das vorzuenthaltende Wissen selbst und können solche Tätigkeiten veranlassen wie Verschweigen, Unterlassen, Vertuschen, Themenwechsel, aber auch Manipulation von Tagesordnungen, Vernichten von Akten, Lügen, Fälschen, Unterschlagen. Zum anderen können die Techniken des Verbergens die Personen der Geheimnisträger selbst betreffen und zu so bekannten Maßnahmen führen wie Verstecken, Maskieren, Tarnen, chirurgische Operation, Selbstentstellung. Nach Simmel nehmen Verbergungstechniken oft die Form einer "aggressiven Defensive" gegenüber dem (oder den) ausgeschlossenen Dritten an.

Die dritte Handlungsart, die nach dem Modell geheimer Dreierkonstellationen entfaltet wird, besteht in dem Versuch des Dritten, das Geheimnis von (A) und (B) zu *enthüllen* (I.3). Um Enthüllungsaktivitäten entfalten zu können, müssen drei Voraussetzungen erfüllt sein; der Dritte muß ein Minimum an Wissen besitzen: Erstens muß er wissen, *daß* er von Wissen ausgeschlossen wurde; zweitens, *wer* ihm Wissen vorenthält und drittens muß er überhaupt daran interessiert sein, das vor ihm verborgene Wissen zu enthüllen.

Während die ersten beiden Voraussetzungen oft erst in einem langwierigen Prozeß erfüllbar sind, der zum eigentlichen Drehpunkt der Interaktionsdynamik werden kann, ist nach Simmel die letzte Voraussetzung am ehesten zu erfüllen. Oft reicht bereits die Kenntnis vom Ausschluß von der Teilhabe des exklusiven Gutes Wissen aus, um die Motivation zur Enthüllung zu wecken. Dabei kann der Dritte einer "typischen Irrung" unterliegen, indem er annimmt: "alles Geheimnisvolle ist etwas Wesentliches und Bedeutsames" (Simmel 1968, 274). Unbekanntes und Unzugängliches erfährt in der Phantasie des ausgeschlossenen Dritten eine Wertsteigerung und Aufmerksamkeitsbetonung, "die die offenbarte Wirklichkeit meistens nicht gewonnen hätte" (Simmel 1968, 274). Eben diese – oft maßlose – Wertsteigerung mobilisiert ausgeschlossene Dritte zu Enthüllungsaktivitäten, die mehr oder minder aufwendig sein können; von so laienhaften und alltäglichen Enthüllungsversuchen wie "das Horchen an verschlossenen Türen und das Hineinschielen auf fremde Briefe", "das gierige, spionierende Auffangen jedes unbedachten Wortes, die bohrende Reflexion: was dieser Tonfall wohl zu bedeuten habe, wozu jene Äußerungen sich kombinieren ließen, was das Erröten bei der Nennung eines bestimmten Namens wohl verrate. . ." (Simmel 1968, 267), bis zu systematischen Techniken professionalisierter und "routinisierter" Enthüller, wie Journalisten, Spione, Detektive, Fahndungsbeamte, Sozialwissenschaftler und Computer (Marx and Reichmann 1984). Der "aggressiven Defensive" der Geheimnisträger steht die "aggressive Offensive" der Enthüller gegenüber – eine Frontstellung, die dann ins Ungleichgewicht gerät, wenn Professionalität und Routine ungleich auf beide Seiten verteilt sind.

## II. Geheimnis als Institution

Der zweite Aspekt des interaktionistischen Paradigma von Geheimnis betrifft die Frage, in welchem Ausmaß Geheimhaltung *institutionalisiert* ist. Verbergen von Wissen kann Positionsinhabern als Pflicht vorgeschrieben sein, deren Einhaltung durch Sanktionen unterschiedlicher Art und unterschiedlichen Ausmaßes sozial kontrolliert wird. Institutionalisierte Geheimnisse sorgen dafür, daß die Kommunikation von Wissen nicht der Idiosynkrasie der jeweiligen Positionsinhaber überlassen bleibt. Vielmehr wird etwa von dem Sozialbeamten, Priester, Arzt erwartet, daß sie als Positionsinhaber ihrer professionellen Pflicht genügen und das ihnen von ihren Klienten anvertraute Wissen Dritten gegenüber verschweigen. Von den Klienten wird andererseits erwartet, daß sie so viel Wissen mitteilen, wie zur Herstellung der von ihnen gewünschten Leistung notwendig ist: Um vom Beamten Kindergeld, vom Priester Gewissenserleichterung, vom Arzt Heilung zu erhalten, muß der Klient (B) Wissen über seine Person bereitstellen. Da (A) in der Regel definiert, wie viel Wissen er zur Herstellung der Leistung benötigt, befindet sich (B) in einer diffusen Abhängigkeit von (A). Er wird sich auf diesen asymmetrischen Tausch nur dann einlassen, wenn er von der Erwartung ausgehen kann, daß (A) seine Schweigepflicht einhält. Die Institutionalisierung von Geheimhaltung ist funktionale Voraussetzung für die Realisierung von Zielen zentraler gesellschaftlicher Einrichtungen (Simmel 1968, 273; Merton 1968, 398f; Popitz 1968). Hält sich auch nur ein Akteur in dieser Triade nicht an die ihm positionell vorgeschriebenen Normen, gerät die betreffende gesellschaftliche Einrichtung ins Wanken: Bricht (A) seine Schweigepflicht, stellt (B) kein Wissen mehr bereit; verweigert (B) die von (A) geforderten Angaben zu seiner Person, stellt (A) die betreffende Leistung nicht her; und ist der Dritte mit Enthüllungsoffensiven erfolgreich, bricht die Tauschbeziehung zwischen (A) und (B) zusammen. Das Institut der Volkszählung ist in der Bundesrepublik aufgrund einer derartigen in sich verzahnten Beziehungskette von nicht gewährleisteter Geheimhaltung, Aussageverweigerung, Leistungsausfall zumindest in der bislang praktizierten Form zusammengebrochen. Gesellschaftliche Einrichtungen, die ihre Ziele nur auf der Basis von Geheimhaltung erreichen können, müssen sich nicht nur effektive Mechanismen zur Kontrolle von Normenverstößen einfallen lassen, sondern vor allem Mechanismen zur Schöpfung von Vertrauen und Herstellung von Sinn.

## III. Geheimnis als Organisation

Von Geheimnis als Institution ist Geheimnis als *Organisation* zu unterscheiden. Schließen sich beispielsweise schwedische Bischöfe einem Freimaurerorden an; bilden Parteifunktionäre Tarnorganisationen; folgen Stu-

denten geheimen religiösen Sekten oder politischen Geheimbünden, so tun sie dies freiwillig. Weder dem schwedischen Bischof, noch dem Parteifunktionär, noch dem Studenten ist es positionell vorgeschrieben, sich einer dieser erwähnten geheimen Organisationen anzuschließen. Im Gegenteil: Sie geraten typischerweise aufgrund ihrer Mitgliedschaft in Geheimorganisationen in Konflikt mit anderen Vorschriften, die an ihrer Position haften. Von Bischöfen wird erwartet, daß sie vor ihrer Gemeinde keine Geheimnisse haben; Parteifunktionäre müssen Spendengelder öffentlich ausweisen; Studenten wird angesonnen, ihre religiösen und politischen Erlösungssehnsüchte innerhalb der etablierten gesellschaftlichen Einrichtungen zu befriedigen. Die sozialen Konflikte, die in diesen Fällen entstehen, sind je anders gelagert; daher scheint es sinnvoll und wichtig, analytisch zwei Typen von Geheimorganisationen zu unterscheiden, primäre und sekundäre.

In *primären* geheimen Organisationen "ist das Geheimnis soziologischer Selbstzweck, es handelt sich um Erkenntnisse, die nicht in die Menge dringen sollen; die Wissenden bilden eine Gemeinschaft, um sich gegenseitig die Geheimhaltung zu garantieren" (Simmel 1968, 289). Die Mitglieder primärer geheimer Organisationen halten Wissen absichtlich und bewußt exklusiv, weil nach ihrer Einschätzung sonst das Organisationsziel nicht realisiert werden könnte. So begründet etwa ein prominenter schwedischer Bischof die Geheimhaltung der Rituale in Freimaurerorden mit der besonderen pädagogischen Wirkung geheimer Praktiken für die zu initiierenden Mitglieder: "Ein Text, den man zum ersten Mal hört, hat eine ganz andere Wirkung und löst ein ganz anderes Erlebnis aus als ein Text, den man von vornherein kennt." (Biskopen i Växjö, Sven Lindegard, skrivelse till ansvarsnämnden för biskopar, 4.3.1985). Primäre geheime Organisationen müssen danach streben, eine bestimmte Organisationsgröße nicht zu überschreiten. Je größer die Organisation, desto stärker wird die Realisierung ihres Organisationsziels gefährdet. Typischerweise geraten numerisch limitierte Geheimorganisationen in ein Dilemma: Je kleiner die noch verträgliche Mitgliederzahl für die Zielerreichung ist, desto größer werden die Probleme ihrer materiellen Versorgung und der Aufrechterhaltung ihrer Autonomie. Dieses Dilemma lösen Geheimnisorganisationen oft dadurch, daß sie ihren wenigen Mitgliedern hohe Beitragszahlungen zumuten. Dadurch erhalten primäre geheime Organisationen oft den Akzent ökonomischer Exklusivität, der nicht immer mit den inhaltlichen Zielen der Organisation in Harmonie gebracht werden kann. Wie ein schweizerischer Freimaurer mitteilte, sah sich sein Orden angesichts seiner zahlenmäßig und finanziell zu schwachen Mitgliederbasis gezwungen, um Mitglieder offen zu werben, ein Mittel, das dem Modellcharakter reiner primärer Geheimorganisationen radikal widerspricht.

*Sekundäre* Geheimorganisationen sind solche, deren Organisationsziele und/oder Mittel von der äußeren Gesellschaft als deviant definiert werden

und die den Weg der Geheimhaltung wählen, um sich der gesellschaftlichen Verfolgung zu entziehen. Weder für die Mafia, noch für die Rote Armee Fraktion, noch für die Muslimbrüder (um nur einige prominente Beispiele sekundärer Geheimorganisationen zu erwähnen) ist Geheimhaltung soziologischer Selbstzweck. Für diese ist sie vielmehr ein – meist aufwendiges – organisatorisches Mittel, um Mitglieder und materielle Ressourcen vor äußerem Zugriff zu schützen und um zu gewährleisten, daß die ursprünglichen Organisationsziele unter der veränderten Bedingung ihrer Kriminalisierung weiter verfolgt werden können (Simmel 1968, 282).

Für primäre Geheimorganisationen ist die Anwendung der soziologischen Technik der Geheimhaltung eine conditio sine qua non für die Zielerreichung; für sekundäre Geheimorganisationen wird die ihnen von außen aufgezwungene (oder nahegelegte) Technik der Geheimhaltung oft zur Voraussetzung für ungewollte und unbemerkte Zielverschiebungen, zur Fehlinvestition von Energien und zur verzerrten Wahrnehmung der Realität von seiten untergetauchter Mitglieder. Friedhelm Neidhardt hat diese Eigendynamik ganz raffiniert am Beispiel der Geschichte der RAF analysiert: Kassiberschreiben, Untertauchen, Wohnungsbeschaffung, materielle Beutezüge etc. wurden unmerklich zur Hauptbeschäftigung und führten die Mitglieder der RAF immer weiter von ihren ursprünglichen Zielen ab (Neidhardt 1981). Die spezifische Gefahr sekundärer Geheimorganisationen besteht darin, daß sich im Laufe ihrer Geschichte die Technik der Geheimhaltung zum Selbstzweck verkehren kann und auf diese Weise die ursprünglichen Organisationsziele und -ressourcen vernichtet werden können. Auch primäre Geheimorganisationen sind der Gefahr von Zielverschiebung, Organisationsdeformation und Selbstdestruktion ausgesetzt. Je mehr Wissen die äußere Gesellschaft produziert, desto schwerer haben es primäre Geheimorganisationen, ihr Wissen exklusiv zu halten. So wurde beispielsweise das ursprünglich von den Freimaurern gehütete Geheimnis der Baukunst durch Verwissenschaftlichung und Popularisierung von Mathematik und Architektonik gelüftet. Die Freimaurer sahen sich angesichts dieser Offenbarung gehalten, die Geheimhaltung auf die formalen Aspekte ihrer Organisation, auf die Rituale, zu verlagern, wenn sie als Geheimorganisation überlebensfähig bleiben wollten. Die Existenz primärer Organisationen ist zusätzlich dann gefährdet, wenn sich das Prinzip der Öffentlichkeit in der äußeren Gesellschaft kulturell und politisch mehr und mehr durchsetzt. Im Zentrum der hitzigen Debatte um die Existenzberechtigung der schwedischen Freimaurerlogen stand die Frage der Legitimität geheimer Organisationen: "Die Freimaurer sind eine geschlossene Gesellschaft, die nicht den heutigen gestiegenen Anforderungen an Offenheit entsprechen" (so etwa Barbro Berglund, Svenska Kyrkans nämnd för mellankyrkliga och ekumeniska förbindelser, Sydsvenska Dagbladet Snällposten, 5.3.1985).

## IV. Die dualistische Struktur von Geheimnis

Simmel geht allgemein davon aus, daß soziale Einheiten erst dadurch an Vitalität gewinnen, wenn in ihnen gleichzeitig gegensätzliche soziale Kräfte zusammenwirken (Simmel 1968, 262; Nedelmann 1980). Dreierkonstellationen haben im Vergleich zu Dyaden und "Mehr-als-Drei"-Gruppierungen durch die schillernde Rolle des Dritten eine besonders hohe Vitalität. Triaden jedoch, die auf der Basis von Geheimnis gebildet werden, geraten dadurch unter Hochspannung, daß sich die Akteure an widersprüchlichen Interessen orientieren, ständig in die Versuchung zum Seitenwechsel und Überlaufen geraten und plötzlichen emotionalen Wechseln ausgesetzt sind. Akteure, die in geheime Interaktionskonstellationen involviert sind, geraten typischerweise unter Streß; Kurzschlußhandlungen und Überraschungsmanöver sind daher wahrscheinlich, insbesondere bei unprofessionalisierten Geheimnisträgern.

Der erste Dualismus in geheimen Dreierkonstellationen besteht in der Freisetzung von sowohl assoziierenden als auch dissoziierenden sozialen Kräften. Auf der einen Seite wird durch das Teilen von Wissen eine Klasse von Besitzern geschaffen, die sich das Recht zur Kontrolle ihres Monopols aneignen und hierdurch intern integriert werden (siehe Wilson 1984, insbs. 17-19). Auf der anderen Seite "legt das Geheimnis eine Schranke zwischen die Menschen" (Simmel 1968, 275), indem es die vom Wissen ausgeschlossenen Dritten als Klasse von Nichtbesitzern konstituiert und in eine antagonistische Position zur Klasse der Wissensmonopolisten bringt. Der "Klassenkampf" zwischen Besitzern und Nichtbesitzern spitzt sich im allgemeinen auf die Kernfrage der Umverteilung von Wissen zu und entbrennt in dem Moment, in dem die deprivierten Dritten Enthüllungsversuche unternehmen und die Geheimnisträger ihr Monopol verteidigen. Jedes Mitglied beider Klassen ist jedoch potentieller Überläufer ins feindliche Lager; daher ist die ursprüngliche Konstellation stets der Gefahr ausgesetzt, sich zu verschieben oder total zu verändern.

Diese intern angelegte Transformationsdynamik ist auf ein zweites dualistisches Merkmal zurückzuführen: Es betrifft die widersprüchlichen Interessen, an die sich die drei in Geheimnisse involvierten Akteure orientieren können.

Was zunächst die Geheimnisträger betrifft, so richten sie ihr Handeln an zwei sich widersprechenden Interessen aus, die beide für sie attraktiv sind. Einerseits sind sie daran interessiert, ihre exklusive Stellung zu erhöhen und ihr "stark betontes Eigentumsgefühl" zu vertiefen (Simmel 1968, 273); dies realisieren sie durch Verteidigung und Expansion ihres Wissensmonopols. Andererseits werden sie zum logischen Gegensatz zur Geheimhaltung hingezogen: zum Verrat. Seine Attraktionen bestehen in dem "verführerischen Anreiz", die soziale Schranke, die sie durch das

Teilen des Geheimnisses zwischen sich selbst und Dritten aufgebaut haben, "durch Ausplaudern oder Beichte zu durchbrechen" (Simmel 275). Bewahren und Verraten des Geheimnisses stellen für die Geheimnisträger je unterschiedliche Gratifikationen in Aussicht: Geheimhaltung wird durch zunehmende Solidarität zwischen (A) und (B) belohnt, Verrat durch die Empfindung der "Lust der Beichte" (Simmel 1968, 275); es steht in der Macht des Verräters, die bestehende Konstellation durch Einsatz geringfügiger Mittel (oft reicht ein einziges Wort) total und überraschend zu verändern. Verräter kommen in den Genuß, Loyalitäten umzupolen: Sie können vorher ausgeschlossene Dritte emotional einbinden und ursprünglich Verbündete fallen lassen. Das Moment des Geheimnisverrates enthält Reize, die denen vergleichbar sind, die der Verschwender von Geld empfindet: "... das mit dem Geldbesitz gegebene Machtgefühl konzentriert sich für die Seele des Verschwenders am vollständigsten und am lustvollsten in dem Augenblick, wo er diese Macht aus Händen gibt."(Simmel 1968, 274). Wenn die Weiterleitung einer wichtigen Mitteilung gelegentlich dadurch zu beschleunigen versucht wird, daß man sie mit dem Siegel der Verschwiegenheit versieht, so bedient man sich eben dieser Erkenntnis.

Da die Geheimhaltung stets durch die Attraktivität des Verrates intern bedroht ist, stiftet sie typischerweise höchst labile Interaktionskonstellationen. Diese intern verursachte Labilität wird zusätzlich durch eine externe Gefahrenquelle verstärkt, durch die Enthüllungsversuche Dritter. Wegen dieser beiden Gefahrenquellen bedürfen geheime Interaktionskonstellationen zusätzlicher sozialer Stabilisierungsmechanismen, sofern sie auf Dauer gestellt werden sollen. Ihre Erörterung verdiente gesonderte Beachtung.

Was den ausgeschlossenen Dritten betrifft, so ist auch dieser in seinen Handlungen durchaus nicht eindeutig ausgerichtet. Seine Enthüllungsabsichten sind beispielsweise dann korrumpierbar, wenn ihn die Geheimnisträger am Genuß des exklusiven Gutes Wissen teilhaben lassen. Wie die meisten Deprivierten, wird auch er dazu neigen, seine "aggressive Offensive" in dem Maße zu mildern, in dem ihm durch das Überlaufen ins feindliche Lager der Geheimnisträger Macht zuwächst. In die Klasse der Geheimnisträger eingeschlossene Dritte senken aber tendenziell die Wahrscheinlichkeit für das Weiterbestehen des Geheimnisses. Je größer die Gruppe der Geheimnisträger wird, desto größer werden die Kontrollprobleme bei gleichzeitigem Sinken des Wertes des Geheimnisses. Schon deshalb werden sich die Geheimnisträger nur begrenzt der Technik der Einweihung ausgeschlossener Dritter in ihr Geheimnis bedienen können, um die aggressiven Enthüllungsoffensiven Wissensdeprivierter zu pazifizieren. Je öfter sie sich dieses Mittels bedienen, desto stärker tragen sie selbst zur Destruktion ihres Geheimnisses bei.

## V. Das psychische Leben von Geheimnis

Soziales Handeln löst bei den wechselwirkenden Akteuren bestimmte Arten des Erlebens aus (Nedelmann 1985). Diese allgemeine Erkenntnis Simmels gilt auch für das Geheimnis als Interaktionsform. Individuen, die in geheime Triaden involviert sind, geraten in typische emotionale Verfassungen.

Das Gefühlsleben der Geheimnisträger ist höchst widersprüchlich und voller Spannungen. Ihr ohnehin "stark betontes Eigentumsgefühl" wird durch das Bewußtsein gesteigert, daß andere von diesem Besitz ausgeschlossen sind. Der Stolz auf die Ausnahmestellung ("Ich weiß doch was, was du nicht weißt", Simmel 1968, 274) wird typischerweise zum formalen Mittel, um gegenüber ausgeschlossenen Dritten zu prahlen und sie zu deklassieren. Beamte bedienen sich vorzugsweise dieses Mittels gegenüber lästigen Klienten; Naturwissenschaftler und Politiker gegenüber protestierenden Kernkraftgegnern und Umweltschützern; Pressesprecher gegenüber neugierigen Journalisten; Vertreter von Regierungspartei(en) angesichts bohrender Fragen von Vertretern der Oppositionpartei(en). Diesem stark betonten Eigentumsgefühl, das mit Geheimhaltung verbunden ist, steht das bereits erwähnte Gefühl der "Lust zur Beichte" gegenüber, der Machtgenuß, der mit dem Moment des Verrates maximiert und gleichzeitig konsumiert wird. Bereits tropfenweise dosierte Informationsflüsse, Andeutungen und Geheimtips mögen für viele Geheimnisträger in ihrem Alltagsverkehr mit unwissenden Dritten Lustgewinne einbringen. Sind Machtgefühle jedoch durch sich widersprechende Handlungen, durch Verbergen *und* Offenbaren, zu befriedigen, dann wird die Berechenbarkeit des künftigen Handelns von Geheimnisträgern schwer, insbesondere von solchen, die machtanfällig und -süchtig sind. Geheimnisse ziehen Persönlichkeiten vom Typus des "excessive ego" an (Wilson 1984, 18); deren Handeln ist bekanntlich häufig abwegig.

Diese prinzipielle Schwierigkeit der Berechenbarkeit des Handelns der Geheimnisträger trägt zur Herauspräparierung des wohl wichtigsten Merkmals des psychischen Lebens in Geheimnissen bei: dem anhaltenden Gefühl der Unsicherheit. Durch die stets präsente Gefahr der internen und externen Bedrohung ihres Geheimnisses müssen deren Träger immer auf "Schicksalswendungen und Überraschungen" gefaßt sein (Simmel 1968, 275). Bekanntlich ist die Überlebenschance solcher Gruppen gering, deren Mitglieder keine gesicherten Hypothesen über das künftige Handeln anderer aufstellen können. Diese Gruppierungen sind funktional auf eine Emotion angewiesen, mit der sie ihre Unsicherheit überwinden können: Vertrauen. Nach Simmel ist Vertrauen eine "Hypothese künftigen Verhaltens", mit deren Hilfe der "mittlere Zustand zwischen Wissen und Nichtwissen um den Menschen" (eben: Unsicherheit) in Richtung Wissen verlagert werden kann (Simmel 1968, 263). Vertrauen ist eine notwendige

Emotion, um "praktisches Handeln" (ebenda) in geheimen Organisationen und Institutionen zu ermöglichen. Der soziologische Fachausdruck der "greedy institution" ließe sich vorteilhaft zur Charakterisierung geheimer Interaktionskonstellationen verwenden: Sie sind eine nach Vertrauen gierige Einrichtung und neigen in ihrer Gefräßigkeit dazu, Individuen in ihrer gesamten Persönlichkeit in die geheime Konstellation einbeziehen zu wollen.

Je mehr geheime Interaktionskonstellationen jedoch dazu tendieren, zu totalen Einrichtungen zu werden, desto größer ist die Gefahr,daß ihnen die Kontrolle des Privatlebens und sonstiger Aktivitäten der ihnen angehörenden Gesamtpersönlichkeiten entgleitet. In diese Gefahr begeben sich geheime Einrichtungen auch dann, wenn ihre Gier auch das Geheimwissen selbst erfaßt: Geheimdienstler, Sozialbeamte, Statistiker, Journalisten, Sozialwissenschaftler und andere professionelle und "routinisierte" Beschaffer persönlicher Daten können typischerweise nur selten ihre Neugier nach immer mehr Wissen beherrschen und leisten damit dem Mißbrauch von Daten Vorschub.

Geheime Gesellschaften begegnen diesen Gefahren in der Regel durch Ausübung "absoluter Herrschaft über die Mitglieder" (Simmel 1968, 299). Sie entwickeln sich zum Typus einer autoritären, zentralistisch und hierarchisch strukturierten Organisation (Simmel 1968, 297-304; Neidhardt 1981).

Das für geheime Triaden typische Gefühlsleben und die durch es verursachte interne Spannung ist erst dann vollständig begreifbar, wenn man die Emotionen des ausgeschlossenen Dritten (C) mit in die Analyse einbezieht: Dieser ist hauptsächlich von Eifersucht beherrscht. Aus dieser Emotion wächst geheimen Interaktionskonstellationen eine Handlungsenergie zu, die erst zur vollen Entfaltung der für sie typischen destruktiven Eigendynamik führt. Bereits die formale Tatsache des Ausschlusses von Wissen weckt im allgemeinen Eifersucht auf die Wissenden. Vermutet der ausgeschlossene Dritte zusätzlich – zu Recht oder zu Unrecht –, daß das ihm vorenthaltene Wissen nachteilig für ihn ist, wird er motiviert genug sein, um die mit Enthüllungsaktivitäten unvermeidlich verbundenen Kosten zu tragen. Simmel hat in verschiedenen Zusammenhängen analysiert, inwiefern Eifersucht eine sozial destruktive Emotion ist. (Das ist an anderer Stelle nachzulesen: Nedelmann 1983, 198-203). Der eifersüchtige Dritte erhebt einen Anspruch darauf, Mitbesitzer des exklusiven Gutes Wissen zu werden; oft muß er sich diesen Anspruch erst politisch erkämpfen und durchsetzen. Dem Prototyp des eifersüchtigen Enthüllers gehen die Energien erst im Moment der Destruktion der geheimen Interaktionskonstellation aus. Ebenso wie der auf Liebe eifersüchtige betrogene Ehemann es eher vorzieht, die Liebe zur Ehefrau zu zerstören, als vorzeitig seinen Anspruch auf Liebe fallen zu lassen, ebenso wird es auch der vom Geheimnis ausgeschlossene Dritte eher vorziehen, das Geheimnis

selbst zu zerstören, als die Waffen gegen die Klasse der Wissensmonopolisten zu früh zu strecken. Zerstören ist in den meisten Fällen auch technisch einfacher, als in den Besitz des Geheimwissens zu gelangen. Auch die Geheimnisträger werden es in vielen Fällen vorziehen, ihr Geheimwissen zu zerstören, bevor es Dritten in die Hände fällt. Wissen ist Macht; durch Zerstören von Wissen wird Macht entzogen – nach dieser bekannten Formel werden beide Lager handeln.

Diese erwähnten Emotionen, Machtgefühle der Wissensmonopolisten und Deprivationsgefühle der Nichtwissenden; Lust des Verbergens und Faszination des Verrates; Unsicherheit und Vertrauen; Besitzerstolz und Eifersucht verleihen in ihrer Kombination geheimen Dreierkonstellationen jene emotionale Attraktivität, die mit erklären mag, weshalb Geheimnis auch dann noch als Interaktions*form* weiterlebt, wenn das Wissen selbst offenbart worden ist. Das Überleben der Freimaurer ist hierfür nur ein Beispiel; die Entstehung neuer freiwilliger Geheimorganisationen, in deren Mittelpunkt die Beherrschung geheimer Rituale steht, ein anderes Beispiel für die noch heute vorherrschende oder wiedererlangte emotionale Attraktivität von Geheimnis als Interaktionsform.

## Bibliographie


Guggenberger, Bernd (1985): Die nackte Wahrheit ist nicht immer das Wahre. In: *Frankfurter Allgemeine Zeitung*, 3.8.1985.

Marx, Gary T. and Reichman, Nancy (1984): Routinizing the Discovery of Secrets: Computers as Informants. In: *American Behavioral Scientist*. vol. 27, Number 4, March/April 1984, 423-452.

Merton, Robert K. (1968): Mechanisms of Observability of Norms and of Role-Performance. In: ders.: *Social Theory and Social Structure*. New York 1968, 395-407.

Neidhardt, Friedhelm (1981): Über Zufall, Eigendynamik und Institutionalisierbarkeit absurder Prozesse. Notizen am Beispiel einer terroristischen Gruppe. In: von Alemann, Heine und Thurn, Peter (Hrg.): Soziologie in weltbürgerlicher Absicht. *Festschrift für René König*. Opladen 1981.

Nedelmann, Birgitta (1983): Strukturprinzipien der soziologischen Denkweise Georg Simmels. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, Jg. 32, 559-573.

Nedelmann, Birgitta (1983): Georg Simmel – Emotion und Wechselwirkung in intimen Gruppen. In: *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, Sonderheft 25, 1983, 174-209.

Nedelmann Birgitta (1985): Action and Experience in Georg Simmel's Sociology, Vortrag gehalten an der Universita` degli Studi di Genova, 16.4.1985.

Popitz, Heinrich (1968): Über die Präventivwirkung des Nichtwissens. Tübingen 1968.

Sennett, Richard (1985): Verfall und Ende des öffentlichen Lebens. Die Tyrannei der Intimität. Frankfurt a.M. und Hamburg 1985.

Sichtermann, Barbara (1985): Ende des öffentlichen Lebens. Zu R. Sennett: Die Tyrannei der Intimität. In: *Ästhetik und Kommunikation*. Heft 57/58, Jahrgang 15, 1985, Intimität, 21-28.

Sievers, Burkhard (1974): Geheimnis und Geheimhaltung in sozialen Systemen. Opladen 1974.
Simmel, Georg (1968): Das Geheimnis und die geheime Gesellschaft. In: ders.: *Soziologie*. Berlin 1968, 256-304.
Wilson, Des (1984): Information is Power: The Causes of Secrecy. In: ders. (ed.) *The Secrets File*. London 1984, 13-21.

# Un schéma d'analyse comparative des entreprises de presse

JEAN G. PADIOLEAU
Paris/Florence

L'analyse économique surprend les sociologues ou les politologues qui d'aventure s'interrogent sur la nature des entreprises de presse.[1] Lorsque, faute de mieux, ces derniers soulignent la nature "hybride" ou le caractère "hétérogène"[2] des organisations productrices de nouvelles, l'économiste perçoit, au-delà des particularités retenant l'attention des spécialistes, des éléments communs à toutes les firmes.

Au regard de l'économie, une station de radio ou un quotidien combine des facteurs de production, leur organisation agence, entre autres choses, capitaux et force de travail, jour après jour ils fabriquent des "produits". Sous cet angle, les firmes de presse ressemblent à des systèmes de mobilisation, d'acquisition et de transformation de ressources.[3] Ainsi un secteur comme celui du commerce fournit-il des ressources monétaires à la presse, lesquelles, en échange de services publicitaires, permettent aux mass-media de fabriquer des biens relativement bon marché; de son côté, la société politique est riche en ressources d'informations qui, à la suite d'un façonnage journalistique, se transforment en nouvelles. Les organes de presse sont donc des "systèmes ouverts" régis par des procédures d'adaptation *interne* et *externe* aux milieux[4] dans lesquels ils oeuvrent. Les produits répondent avec plus ou moins de bonheur à la demande des marchés; les modes de fonctionnement et les structures des firmes s'efforcent de "coller" aux caractéristiques des environnements; la nécessité d'atteindre par exemple des "cibles" d'audiences définies conduit les quotidiens à créer des départements publicitaires ou de diffusion respectant les profils démographiques.[5] Ces phénomènes d'adaptation recouvrent de même des procédures d'organisation du travail journalistique. Outre-atlantique, à la division du gouvernement Fédéral en agences et en départements, correspond un mode d'allocation des reporters dénommé "beat"[6]; cette procédure d'allocation géographico-bureaucratique des reporters facilite la gestion du personnel des salles de rédaction: le procédé du "beat" permet aussi aux rédacteurs en chef de ne pas être pris au dépourvu. Ou bien encore: des journaux se risquent dans des activités d'adaptation "externe" quand

ils créent de nouveaux produits ou lorsqu'ils choisissent des politiques de diversification ou d'acquisition d'entreprises concurrentes.[7]

Quelles que soient ses modalités, l'objet de l'analyse économique est de montrer comment, d'une manière ou d'une autre, les partenaires d'une entreprise de presse effectuent des calculs s'inspirant de critères d'évaluation utilitaire plus ou moins flous, qui leur permettent d'établir des rapports entre les avantages et les inconvénients des actions.

De l'approche économique, retenons quelques idées-force nécessaires à l'élaboration d'un schéma d'analyse des entreprises de presse productrices de nouvelles et fonctionnant dans le cadre des sociétés libérales oligarchiques-pluralistes.[8] La notion banale de la firme comme "système ouvert" rappelle que les mass-media sont engagés:

(i) dans des *opérations de transformation* de *ressources hétérogènes*; pour ce faire, par suite de cette *dépendance* à l'égard de milieux

(ii) les organes de presse établissent des *liens* avec une *variété d'environnements*, qu'il s'agisse du secteur marchand, de la société politico-étatique, ou des publics;

(iii) aux contacts de ces milieux, les mass-media s'insèrent dans des *rapports-types* avec les acteurs sociaux constitutifs des environnements;

(iv) ces relations *varient en nature*, et tout spécialement les principes de rectitude ou d'efficacité qui les guident n'entretiennent pas nécessairement des rapports directs de similitude et de correspondance

(v) qui plus est, dans l'action quotidienne ces rapports *s'interpénètrent.*

Justement, ces traits fondamentaux des moyens d'information nous ont incité à prendre le risque d'une analyse d'inspiration parsonienne de l'entreprise de presse.[9] Cette référence à Parsons ne surprend pas, bien que les travaux sur les mass media s'y réfèrent peu[10], puisque pour l'auteur de *The Structure of Social Action* l'action individuelle on collective s'inscrit dans des contextes. Selon Parsons l'action sociale résulte des rapports d'interdépendance et de complémentarité qu'elle entretient avec des environnements constitués d'acteurs et de ressources diverses; plus encore, l'oeuvre parsonienne, attentive à la circulation de ressources hétérogènes, s'intéresse à leur transfert d'un milieu à l'autre et aux procédures de coordination entre les acteurs. Dés lors, le cadre d'analyse présenté dans les pages suivantes se dispense de définir a-priori les caractéristiques intrinsèques des entreprises de presse: ici, les singularités et les constantes des firmes dépendent des caractéristiques des rapports – types que les mass-media tissent avec les milieux environnants, ou pour mieux dire, les organes de presse ne sont pas considérés comme des "objets", il faut y voir des *configurations* plus ou moins stables de flux de ressources permettant aux firmes d'être et de durer.[11] Dans cette perspective, l'étude ci-jointe dresse en premier lieu un inventaire des rapports-types faisant ressortir quelques – unes de leurs caractéristiques sans pour autant prétendre qu'ils correspondent en tous points aux liaisons concrètes (I); l'analyse explore

ensuite les conditions de "l'institutionnalisation" des rapports-types (II); l'enquête précise aussi quelques-unes des propriétés du système de rapports (III) tandis que la conclusion illustre l'intérêt de ce type de démarche pour des recherches comparatives et théoriques.

## Un inventaire des rapports-types

Le rapport qui s'établit entre les mass-media et les publics est loin d'être simple. Le schéma traditionnel, enveloppé dans des interprétations behavioristes mettant face à face des émetteurs et des récepteurs, est par trop empreint d'une vision mécanique et limitée du rapport[12]; les interprétations fonctionnalistes, bien qu'elles tiennent compte des besoins ou des usages des consommateurs, demeurent dans la veine behavioriste et elles réussissent fort peu à mettre en forme la nature interactive du rapport entre les mass media et les publics. C'est précisément l'intérêt de la réflexion d'O.Burgelin[13] qui interprète cette liaison en termes de "marché" – c'est-à-dire: la rencontre d'une offre et d'une demande perçues comme réciproquement bénéfiques par les partenaires. Selon le schéma du marché, les flux d'offre et de demande des messages entrent en rapport par suite de l'ouvrage d'un lien-type entre les protagonistes, correspondant à une transaction au cours de laquelle le souci de réussir des opérations bénéficiaires guide les partenaires. Les publics assurent des rôles de lecteurs ou d'auditeurs lorsqu'ils perçoivent dans ces rôles des activités en accord avec leurs préférences; pour leur part, les industriels de la presse, soucieux de performances, tentent de satisfaire les consommateurs par la fourniture de biens appropriés. Ainsi donc la formule du marché met-elle en lumière la nature fondamentalement interactive du lien unissant les publics aux moyens d'information. Comme le dit fort bien O.Burgelin:

> "En termes sociologiques, parler de demande c'est considérer le public non comme une audience mais comme un ensemble *d'acteurs* (. . .). Car se référer à la notion de marché c'est introduire l'idée d'une action réciproque de l'offre et de la demande".[14]

La figure du marché rappelle aussi que les producteurs et les consommateurs, dans les systèmes de mass-media, possèdent des *capacités stratégiques* (i.e. des aptitudes variables à agir et à réagir).[15]

Bien que la formule du marché décrive les grands traits morphologiques de l'interaction entre la presse et les publics, peut-on se contenter d'y voir une simple transaction? Certes, les enquêtes d'audiences montrent que les conduites des consommateurs visent à satisfaire des besoins d'information, de détente ou de culture; d'une manière ou d'une autre, les lecteurs ou les auditeurs évaluent les satisfactions effectivement fournies. Mais, avec non

moins de force, ces derniers acceptent de tenir des rôles de récepteurs dans l'espoir que l'interaction avec les producteurs satisfasse des *attentes* dont certaines dépassent un cadre stricto-sensu fonctionnel-utilitaire.

Appliquons par exemple cette formulation en termes d'attentes aux nouvelles politiques; dans ce cas, les citoyens acceptent au préalable comme allant de soi, ou avec plus d'à-propos comme légitime, la division sociale du travail présente dans l'ouvrage journalistique; au cours de l'interaction entre la presse et les publics, un produit comme les "news" est légitime sous réserve que des "attentes réciproques" rendent ce bien acceptable et convenable, intelligible et significatif aux partenaires. La nouvelle n'est pas tout simplement utile, des données, outre la valeur d'usage, "donnent du sens" comme disent les linguistes à la communication.[16] En effet, dès que l'on complète le schéma du marché par une formulation plus générale en termes d'attentes, les principaux *modus operandi* constitutifs de l'interaction entre la presse et les publics se précisent.

La rencontre des flux d'offre et de demande nécessite en premier lieu des (i) *phénomènes de dépendance réciproque* entre les fabricants et les consommateurs de messages. Les peuplades d'Amazonie n'éprouvent pas le besoin d'une communication de masse puisqu'ils peuvent assurer par eux-mêmes la collecte ou la diffusion des nouvelles; à l'opposé, dans l'essor des phénomènes d'interdépendance, par suite de l'extension de la démocratie politique et de la division sociale du travail, réside l'une des causes premières de l'envol de la grande presse urbaine et nationale – ce qui n'échappait pas à un auteur comme R.E.Park.[17] Néanmoins, la dépendance fonctionnelle ne suffit pas à créer, et plus encore à asseoir, l'interaction entre la presse et les audiences: les protagonistes doivent aussi partager (ii) un fonds commun de principes et de valeurs, ou, en langage parsonien, des *solidarités*, ne serait-ce que pour apprécier les inconvénients ou les avantages des phénomènes de dépendance.

Considérons sous cet angle l'un des biens de l'industrie de la presse comme les nouvelles politiques. La valeur de ce type d'information résulte de l'assentiment des citoyens à des principes de l'idéologie démocratique-libérale comme celui de la surveillance des élites politiques par les gouvernés. Tocqueville ne s'y trompait pas, "les journaux ne se multiplient pas seulement en suivant le bon marché disait-il, mais suivant le besoin plus ou moins répété qu'un grand nombre d'hommes ont de communiquer ensemble et d'agir en commun".[18] Traduits dans un langage d'"attentes réciproques", les phénomènes dégagés par Tocqueville – "communiquer" et "agir en commun" – exigent que la division sociale du travail journalistique englobe des (iii) *sentiments de confiance* (iiii) présents dans des zones d'activités délimitant les sphères et les attributs des *compétences* respectives des partenaires de l'interaction media-publics.

Dès lors, si le langage des attentes enrichit notre compréhension du

rapport autre les mess media et les publics, de plus, il en dévoile un ressort essentiel: la présence d'engagements réciproques (*commitment*) d'*Ego* et d'Autrui, producteurs d'états généraux de *crédit* – et singulièrement celui des récepteurs vis-à-vis des producteurs de messages. Dans ces circonstances, comment qualifier le rapport entre la presse et les publics? Celui-ci ne se laisse pas réduire à la formule du marché, on ne peut le confondre avec l'obligation morale, il se distingue à l'évidence du pouvoir puisqu'il ne recourt ni à la force ni à la menace. Néanmoins, les traits du dit rapport ne nous semblent pas étrangers, puisque ses caractéristiques générales correspondent au phénomène de la persuasion – c'est-à-dire selon la terminologie parsonienne à la présence d'un rapport *d'influence*. Dans une relation d'influence, *Ego* ne cherche pas à obliger ou à menacer *Alter*. *Ego* tente de persuader *Autrui* que tel ou tel acte est favorable à son bien-être *et* à celui d'une communauté à laquelle sont sensés appartenir *Ego* et *Alter*. Précisons: dans le rapport d'influence entre les publics et les mass-media tel que nous le concevons, les protagonistes acceptent par *hypothèse* que la communication est en *principe* acceptable et convenable par suite d'accords fondamentaux sur un *cadre* d'attentes réciproques, "diffuses" – la confiance, la compétence, etc – et dont il faut souligner la nature éminemment contingente.

Ce rapport d'influence avec les publics se distingue des liens noués par les mass media avec l'environnement du secteur économique. Ces relations s'inscrivent dans un contexte d'attentes explicitement utilitaires, ce sont des *changes économiques*.[19] Des industriels vendent de la pâte à papier, les entreprises engagent des salariés, les journaux offrent aux annonceurs des espaces publicitaires moyennant recettes et les publicitaires signent des accords stipulant avec précision les engagements réciproques des protagonistes. Dans un rapport d'échange économique, à la différence de l'influence, où prédomine un cadre d'attentes générales, "diffuses", et par conséquent plus ou moins incertaines, un *contrat* formalise les flux de ressources utilitaires.

Cette situation contractuelle diffère avec les attentes constitutives des liens-types qui relient les moyens d'information et la société politico-étatique. Dans ce contexte, les protagonistes, d'une part, entretiennent des rapports d'étroite dépendance, d'autre part, ils ressentent des sentiments élevés d'incertitude. Les élites politiques, soucieuses de succès électoraux ou de popularité, perçoivent dans les mass-media un canal d'accès privilégié aux opinions publiques, la presse n'ignore pas non plus sa dépendance à l'égard du monde politico-étatique. Les élites constituent l'une des principales sources de nouvelles; le législateur réglemente les activités de la presse, voire même l'Etat participe à son bien-être économique. Quoique cette dépendance soit à des degrés divers de l'ordre de l'inéluctable, la presse n'accepte pas en toutes circonstances une soumission trop forte aux desiderata de l'univers politique. C'est dire combien,

dans les sociétés pluralistes — oligarchiques, la relation mass-media/société politique a toutes les chances d'être foncièrement instable; dès lors, les partenaires aspirent, chacun à leur manière, à des rapports réduisant les zones d'ambiguité et d'incertitude.

Ainsi, les élites politiques ne rechignent-elles pas, selon des degrés variables, à faire usage de *mécanismes de contrôle*. Ces liens de contrôle recouvrent soit (i) *des dispositions formelles contraignantes* — des textes réglementent le travail des journalistes, des lois définissent le statut des établissements ou organisent le choix des dirigeants[20], soit (ii) *des procédures discrétionnaires d'allocation des ressources* nécessaires à la marche des mass-media.[21] Dans ce dernier cas l'Etat peut s'approprier (a) l'usage de biens (les ondes hertziennes) ou de ressources (le paiement d'une redevance obligatoire) ou encore la puissance publique, des syndicats, des partis politiques perturbent (b) le fonctionnement normal des marchés (subventions, régimes fiscaux particuliers, etc).

Dans l'ensemble, les liens de contrôle régissent plutôt les modalités globales d'existence et de fonctionnement des entreprises, mais ces mécanismes n'assurent pas les élites politiques que les activités quotidiennes des mass-media se dérouleront toujours conformément à leurs espérances. Tel est le cas par exemple du commerce journalier entre les sources de nouvelles et les reporters. Quoique ces relations possèdent des visages multiples et variables selon les contextes, nous pouvons néanmoins les rattacher à la catégorie de *l'échange social* dont parle P.Blau.[22] Dans l'échange social, à la différence des contrats économiques, toutes les clauses, y compris parfois les principales, ne sont pas définies avec précision, et elles ne jouissent pas des avantages de la contrainte légale associée souvent aux mécanismes de contrôle formel. Voici, pour illustrer, la rencontre d'un journaliste et du leader d'un parti politique: ce dernier n'ignore pas les capacités de montage et de façonnage des nouvelles dont disposent les journalistes et les rédacteurs en chef, néanmoins l'homme politique accepte de donner un entretien ou de communiquer une information inédite sous réserve que l'échange se déroule dans des conditions convenables et acceptables délimitant à l'avance les contours probables de la nouvelle. Par suite, les rapports du type de l'échange social présupposent la présence et l'action de sentiments de confiance,l'admission de zones de compétences réciproques qui, dans les faits, se traduisent sous la forme de pratiques ou de conventions afin de policer les échanges entre les milieux politiques et journalistiques, ("off the record", "background","deep-background", etc).[23]

Ces visages de rapports d'échange social illustrent un problème commun à toutes les liaisons-types dont nous venons de faire l'inventaire: *instituer* des relations entre les membres d'un système de communication un tant soit peu stables délimitant les contours possibles et probables des actions ou des réactions émises par les protagonistes.

## L'institutionnalisation des rapports

Le problème de l'institutionnalisation s'avère crucial dans le cas de la relation d'influence avec les publics. Des données pratiques montrent son importance: un rapport d'influence stable facilite les activités des firmes, ou tout du moins il n'entrave pas en principe leurs performances économiques. De fait, pour parler comme A.M. Okun[24], une firme de presse apprécie le soutien de "clientèles" – c'est à dire des groupes d'acheteurs ayant une préférence durable pour la firme fournisseuse et non plus des consommateurs anonymes et interchangeables.

Ces phénomènes d'institutionnalisation se révèlent d'autant plus décisifs étant donné certaines caractéristiques inhérentes au processus de façonnage, de diffusion et de réception des "news". Le flou caractérise plutôt la substance cognitive de ces messages; selon les époques, les pays ou les media, la notion même de nouvelles varie; l'ouvrage journalistique nécessite des procédures de mise en forme qui, d'une manière ou d'une autre, s'apparentent à des activités arbitraires; les métiers du journalisme ne disposent pas d'un corps de règles, de concepts ou de méthodes délimitant des savoir-faire comparables en rigueur avec les disciplines de la médecine ou du droit; bien qu'un grand nombre se définissent comme des "professionnels", les reporters n'en demeurent pas moins souvent, comme tout un chacun des citoyens, soucieux à l'occasion de faire prévaloir des opinions privées ou de poursuivre des intérêts particuliers. Ces traits d'incertitude et d'ambivalence s'avèrent d'autant plus essentiels qu'en règle générale, les produits journalistiques sont d'une part visibles, d'autre part fortement valorisés par des acteurs sociaux comme les élites politiques et qui *de facto* tiennent des rôles de surveillants ("Watchdog"). Par suite la liaison avec les publics ne peut être instituée à n importe quel prix, un impératif de "service" – dans le sens parsonien de "collectively oriented" – est consubstantiel au rapport d'influence.

En effet la division sociale du travail journalistique présuppose la différenciation et la spécialisation des rôles d'émetteurs et de récepteurs. Dans cette configuration hiérarchique où Autrui est un récepteur occupant une position subalterne, la confiance des lecteurs ou des auditeurs est susceptible d'être abusée; les intérêts de tiers peuvent requérir aussi la protection (par exemple, délits de diffamation), voire même les collectivités nationales-étatiques ne sont pas nécessairement indifférentes à la variété des modalités d'expression du rapport d'influence quand elles invoquent les impératifs de la raison d'Etat ou de la discipline partisane pour orienter le secteur des mass-media.

Le risque n'est en effet pas mince de restreindre la notion d'influence à l'usage commun dont font état les travaux sur les mass-media. Les messages, répète-t-on, "influencent" (ou "n'influencent pas") les consommateurs de nouvelles; *Ego* accepte ou non les propos ou les avis d'*Alter* par

suite de qualités attachées à des messages particuliers émis et reçus dans des contextes spécifiques. De ce point de vue, être "influent" est sans aucun doute l'enjeu essentiel des "communications impressives"[25] dont les formes les plus notoires sont celles de la propagande et de la publicité. Bien que cette conception, "réaliste" pourrait-on dire, de l'influence soit connexe de l'interprétation "générative" esquissée précédemment, il convient de les disjoindre.[26] Lorsque l'interprétation "réaliste" s'attache aux caractéristiques particulières des actes de communication, (les traits des messages, les qualités des émetteurs, les dimensions des contextes de réception),[27] une perspective "générative" de l'influence souligne les conditions générales, elle met en valeur les *modus operandi* permettant au rapport de s'établir. En vérité les principales composantes de l'influence décrites ci-dessus (solidarité, compétence, confiance, etc.) équivalent à des *capacités* "diffuses" d'interaction se concrétisant par des règles, par des procédures, par des conventions tacites ou réfléchies qui rendent possibles et probables l'émission et la réception des messages. Toutefois, la reconnaissance de ces conceptions "réaliste" et "générative" de l'influence exclut d'y voir des phénomènes entretenant des liens de stricte indépendance: si la conception "réaliste" de l'influence gage principalement l'efficacité du rapport sur des qualités intrinsèques comme celle du contenu des messages, la relation nécessite l'ouvrage concomitant de répertoires communs généraux qui "donnent du sens" aux éléments constitutifs des messages.

Somme toute, le rapport d'influence interprété dans une perspective "générative" assure les rôles d'un "code" qui structure et contrôle les relations entre *Alter* et *Ego*. Les fonds communs d'attentes permettent à *Ego* (ou à *Autrui*) de prévoir, d'explorer le possible et le probable; en même temps, ce substratum d'attentes rend *Ego* lui même prévisible à *Autrui*. Par contre, selon l'interprétation "réaliste" de l'influence, le contrôle de la communication réside beaucoup moins dans la présence d'un cadre de principes généraux que sous la forme d'une régulation par des normes particulières et souvent explicitement contraignantes, ou par des attributs spécifiques aux messages, aux acteurs et aux situations.

Bien qu'il soit nécessaire de garder à l'esprit le caractère analytique de ces distinctions, certains pays et des moyens d'information privilégient, avec plus ou moins d'ampleur selon les conjonctures historiques, l'une ou l'autre des modalités du rapport d'influence auxquelles vont correspondre des types contrastés de journalisme. A ce propos, distinguons, quitte à emprunter librement au vocabulaire parsonien, des rapports d'influence dans lesquels dominent des éléments *particularistes*. Dans ce cas le rapport d'influence à dominante particulariste "active" en premier des critères affectifs d'appréciation qui prennent le pas sur les considérants cognitifs; l'expression des sentiments personnels, culturels, ou politico-idéologiques prédominent; ensuite *Ego* ou *Alter* évaluent de préférence les objets de la communication par rapport à eux-mêmes ou vis-à-vis de situations, de

groupes de référence ou d'appartenance spécifiques pris dans leurs singularités. Par contre, dans un *rapport d'influence à vocation universaliste*, soucieux d'installer ou d'entretenir des conditions générales d'interaction, les éléments cognitifs prédominent; les acteurs qui s'en réclament aspirent à une neutralité affective, ceux-ci invoquent volontiers un cadre de références permettant d'évaluer les actes de communication d'après des critères généraux qui seraient applicables d'une manière universelle aux acteurs, aux situations, ou aux objets analogues; lorsque le rapport d'influence particulariste valorise plutôt des références à des "communautés" singulières riches de qualités propres, la variante universaliste, en contrepoint, se réclame fréquemment d'une collectivité sociétaire saisie comme un ensemble homogène et de ce fait plus propice à la réception de critères universels d'appréciation des performances.

Concrètement le rapport d'influence à vocation particulariste s'illustre d'une manière visible dans le domaine journalistique sous la forme de jugements et d'opinions valorisés idéologiquement; *Ego* indique de bonne grâce à *Alter*, en faisant appel principalement à des communautés de sentiments ou d'idées, ce qu'il faut penser, dire on faire. Une mise en parallèle des presses française et nord-américaine permet de rendre sensible quelques éléments du contraste universalisme – particulariste. Ainsi, selon P. Albert,

"Le journalisme français a-t-il toujours été plus un journalisme d'expression qu'un journalisme d'observation: il accorde la préférence à la chronique et au commentaire sur le compte-rendu et le reportage. Autant qu'à la présentation des faits, il s'est toujours intéressé à l'exposé des idées, autant qu'à l'analyse des situations. Par là il est fondamentalement différent du journalisme anglo-saxon pour qui la nouvelle a toujours la priorité sur le commentaire".[28]

Ce cadre d'analyse, opposant à dessein d'une manière typique les pôles de l'influence particulariste et universaliste, présente en effet l'avantage d'introduire une conception dynamique et stratégique du rapport d'influence entre la presse et les publics. Certes, au cours du temps, selon les pays, au fil du cycle de vie des entreprises de presse, la place respective des rapports d'influence change, et ces derniers peuvent aussi coexister. Néanmoins, il ne semble pas impossible de dégager quelques-uns des facteurs rendant compte des variations. Il en est ainsi de conditions générales favorables à l'institutionnalisation des rapports d'influence. L'accroissement des phénomènes de dépendance des citoyens vis à vis des moyens d'information augmente les chances d'institutionnalisation du rapport d'influence. Ce n'est donc pas un hasard si au regard de l'histoire, outre les effets conjoints de l'urbanisation et du développement de l'instruction, la grande presse s'est développée depuis la fin du 19eme siècle de manière concomitante avec l'essor de l'économie de marché et l'extension de la démocratie politique. La consommation de masse présuppose la mise en

rapport, via la publicité, des vendeurs et des acheteurs, la démocratie politique étend le contrôle et la surveillance des élites par les citoyens, mais pour ce faire, ceux-ci nécessitent d'être informés. En outre, certaines caractéristiques techniques de la communication moderne créent des liens d'interdépendance inéluctables visibles par exemple dans la nature de bien collectif des ondes hertziennes.

Toutefois, l'on ne saurait trop se garder de voir, dans les effets de ces conditions générales, une trajectoire évolutionniste menant, par exemple, d'un journalisme "particulariste" à des versions "universalistes". Ainsi, quoique les Etats-Unis et la France appartiennent à la même famille de société, les expériences nationales modèlent les configurations du journalisme de l'un et de l'autre pays. Outre des mesures légales plus contraignantes d'exercice de la liberté de la presse, la démocratie française, par comparaison avec les Etats-Unis, semble préférer, au modèle du citoyen actif valorisé par le civisme américain, le type du "sujet" réduit souvent à la condition d'électeur; la présence d'un Etat puissant, autarcique, dépositaire et gardien de l'idéologie d'un intérêt général intouchable, prédispose à des attitudes et à des conduites passives, voire de soumission; cet appareillage politico-étatique, de concert avec la centralisation parisienne et l'homogénéité sociale des élites, favorise peu l'ouverture de la société politique, à la différence du style démocratique en vigueur outre-atlantique. De plus, un pays comme la France découvre après les Etats-Unis une véritable économie de marché et plus encore la modernité de la consommation de masse, par suite, entre autres choses, de la pérennité d'une économie rurale. De fait, la faiblesse des ressources publicitaires a constitué, pour paraphraser P. Albert, une sorte d'atavisme de la presse hexagonale — et ce depuis les premiers jours de la grande presse populaire. Il faut aussi tenir compte du rôle d'événements historiques qui, quoique l'analyse sociologique des mass media les négligent, imprègnent les systémes nationaux d'information. Si l'Amérique du Nord a pu demeurer à l'écart de la guerre des propagandes des années trente ou ne pas subir les malheurs de l'occupation et de la collaboration, à l'inverse, de telles épreuves ont profondément pesé sur les mass-media du continent européen notamment en ce qui regarde le rôle de la puissance publique.

Car, en plus de ces conditions générales, les pouvoirs publics, les entreprises et les journalistes, sinon même les citoyens, disposent d'atouts pour infléchir les phénomènes d'institutionnalisation des rapports d'influence dans un sens particulariste ou universaliste. Ces actions peuvent porter sur l'une des dimensions de l'influence comme celle des solidarités. Ces dernières reposent sur des fonds communs de principes ou de valeurs et sur des procédures, comme celles de la concurrence des idées ou des intérêts, supposées donner une existence concréte aux solidarités. La nature principalement symbolique de celles-ci favorise les entreprises des acteurs sociaux pour les orienter éventuellement dans des sens qui leur

conviennent. De plus, par suite de la précarité inhérente au travail journalistique, les engagements de confiance, de compétence ou de service, nécessaires à la création, des liens de solidarité, possèdent un statut conditionnel invitant les partenaires d'un système de communication à agir. En vérité, cette situation contingente permet à nombre de contrôles formels, comme ceux liés à l'intervention des pouvoirs publics, d'acquérir une légitimité, elle favorise aussi l'essor des phénoménes de "professionnalisation" – lesquels méritent un examen attentif si l'observateur entend se tenir à l'écart des rhétoriques de faire valoir.

H.Becker propose à ce propos une définition ironique de la "professionnalisation" qui répond aux besoins de l'enquête: "les activités ayant la chance de se prévaloir et de bénéficier dans le monde du travail d'aujourd'hui du titre honorifique de profession".[29] Cette vue fait donc l'économie des attributs formels réunis dans les professions libérales les plus notoires (médecine, droit); la formule de Becker suggère l'action de mécanismes d'étiquettage social: les membres d'une sphère d'activités se cataloguent, réclament un label de profession; bon gré mal gré, le corps social leur reconnaît éventuellement ce caractère. Sous cet angle, la "professionnalisation" accompagne les démarches stratégiques d'un groupe social pour pouvoir se flatter auprès de l'opinion d'un statut professionnel assurant des privilèges d'exercice hors du lot commun. Précisément les phénomènes de "professionnalisation" surviennent à l'occasion ou pour prévenir des conflits capables de naître entre les acteurs sociaux par suite d'activités ambigües.[30] Plus la division du travail, comme celle du journalisme, crée des incertitudes – susceptibles de provoquer des dommages externes individuels (diffamation) ou collectifs (par exemple la "désinformation" d'opinions publiques) – entre les acteurs sociaux, plus les chances s'élèvent que d'une manière ou d'une autre les activités seront soumises à des instruments de réglementation formelle ou de régulation sociale.[31]

Ainsi, dans le domaine qui nous préoccupe, la "professionnalisation" prétend-elle délimiter les domaines légitimes d'activités des journalistes et des organes de presse ou de circonscrire les frontières d'un bien symbolique comme les nouvelles.[32] Dans les sociétés anglo-saxonnes, les "news" doivent faire montre de traits qui les distinguent de la communication à finalité commerciale ou partisane. L'illustration la plus manifeste d'un tel mouvement est assurément la formation progressive de l'objectivisme journalistique nord-américain depuis la fin du siècle dernier, dont la trajectoire conduit à des phénomènes voyants de "professionnalisme": constitution d'un codex de procédures standards et routinisées de montage et de façonnage de l'information, promulgation de codes déontologiques, naissance d'écoles professionnelles. Dans ce contexte, les nouvelles se trouvent soumises à des critères de reconnaissance et d'évaluation par des publics – principalement, les collègues, les sources ou les patrons – possédant des ressources de légitimation et de sanction. L'impersonnalité

relative des tâches journalistiques qui résulte de ce professionnalisme permet de contrôler les humeurs, les opinions particulières des reporters, tandis que la standardisation du travail journalistique généralise la valeur d'usage des nouvelles.

Si toutes ces pratiques de "professionnalisation" concourent à rendre sensible la distinction chère aux milieux des media entre "l'amateur" et le "professionnel"[33] qui se veut compétent et digne de confiance, d'autres démarches entreprises par la presse s'efforce d'activer des sentiments de solidarité. La catégorie commerciale des audiences n'est pas en effet celle qu'invoquent sur la place publique les journalistes et les patrons de presse – pour s'en convaincre, il suffit de jeter un oeil sur les débats enveloppant les conflits avec les autorités politiques ou judiciaires. Si dirigeants et reporters ne tiennent pas pour inconvenantes les "approches marketing" qui prescrivent de coller aux "cibles", tout du moins essaient-ils de ne pas s'en flatter outre-mesure quand il s'agit de promouvoir ou de défendre le métier de journaliste. La presse promeut plutôt avec un zèle de vrai croyant une rhétorique de justifications, constituées d'objets, de principes politico-juridiques qui, sous l'emprise du consensus libéral, devrait s'imposer à tous: "The Pursuit of Truth", dans les pays anglo-saxons, "le droit à l'information" en Europe. Ce faisant, tout se passe comme si la presse s'identifiait au Peuple, au Public – bien sûr au sens politique classique, celui de Cicéron par exemple: "Le Public ne se réduit pas à un simple rassemblement, il s'agit d'une communauté d'intérêts et de droits". (*De Republica*, 2/42) Représentante par procuration du Public, la presse s'auréole alors tout naturellement d'un *mandat* collectif qui l'autorise à remplir, au nom du Peuple, des missions. Ce mandat semble légitime parce que comme dit T.Burns, la presse jouirait d'une *licence* par suite de sa "professionnalisation".[34]

De fait, la presse peut tendre vers cette autorité – élément essentiel d'un rapport d'influence universaliste – si elle respecte d'une manière loyale et visible les exigences de la "professionnalisation". Mais cette dernière contribue aussi à faire de la presse un acteur collectif ou avec plus de précision, quitte à faire usage d'une notion familière au droit anglo-saxon un "corporate-actor" – c'est à dire une personne morale collective possédant une identité et une conscience de groupe propres[35]. Le journalisme devient ainsi un *objet social* dont l'existence serait indépendante des sujets – des rédacteurs, bien sûr, mais aussi du public et des citoyens.

Toutefois la réification sociale d'un tel sujet collectif comme la Presse exige pour le moins l'accord sur des règles constitutives du journalisme comme objet social, et en particulier de faire sien une croyance radicale résumée en termes simples par M.Schudson: "la conviction que l'on peut et que l'on doit séparer les faits des valeurs".[36] Sous l'effet de cette croyance fondamentale, l'application des règles de la profession permet aux "news" d'être jugées de l'ordre du "fait". Il faut en prendre bonne note, ce procès

capital de légitimation de l'ouvrage journalistique ne nécessite pas des arguments justificatifs qui allègueraient des valeurs finales comme celle de la vérité: l'obéissance à des règles scelle la légitimité de l'ouvrage professionnel.[37] Ainsi donc, contrairement à des idées reçues, l'orthodoxie du journalisme à l'américaine ne prétend pas aujourd'hui qu'elle imprime "le réel" ou qu'elle s'en remette à un empirisme naïf vivace au début du siècle – "les faits parlent d'eux mêmes". De nos jours l'objectivisme correspond à suivre des *procédures* – tenues pour correctes par la profession – de constat des nouvelles (identification des sources, points de vue des parties concernées, etc). Ces éléments de professionnalisme qui se veulent "comme allant de soi" (*given*) présupposent bien sûr la présence d'un facteur décisif: l'état d'une société à dominante consensuelle où les clivages ne se reproduisent qu'imparfaitement dans des luttes idéologiques et partisanes intenses mais dans laquelle les conflits opposent plutôt les "faits" aux "faits". En vérité seules des sociétés "consensuelles" acceptent sans coup férir la réalité indépassable des "faits" pour définir les nouvelles – comme l'illustre de façon exemplaire et limite ceux des media américaines les plus prestigieux qui valorisent des rapports d'influence universaliste.

> La fonction principale de la presse dans une société de consensus, affirme S. Huntington, est de renforcer les accords fondamentaux et de faciliter le jugement des citoyens et des institutions. Justement, outre – Atlantique, la Presse n'est presque jamais gouvernementale, rarement partisane, elle aspire à représenter la société. Le pouvoir de la presse vient de ce que celle-ci est un produit du consensus libéral américain. Les media tiennent le rôle du clergé dans une société libérale: la presse se veut la gardienne de valeurs suprêmes(. . .). Seules les sociétés "consensuelles" peuvent accepter la révélation de faits indiscutables.[38]

Par contre les sociétés politiquement et idéologiquement fragmentées comme celles de l'Europe du Sud se montrent plus rétives aux phénomènes de "professionnalisation".[39] Plus les accords sur le désirable individuel ou collectif et sur les moyens d'y parvenir sont faibles, plus la lutte entre des partis rivaux ou ennemis s'avère intense, plus les chances de rapports d'influence à vocation particulariste s'élèvent. Les clivages présents dans de telles sociétés correspondent à des groupes sociaux, culturels et politiques, favorables à des communications respectant leurs singularités; dans de tels contextes, la communication journalistique tout à la fois exprime et participe à l'identification réciproque des groupes et des conflits qui les particularisent. En France ou en Italie, les liens des lecteurs communistes avec l'*Humanité* ou avec l'*Unita* s'appuient sur des croyances et des valeurs spécifiques créateurs de sentiments réciproques de solidarité, de confiance ou de compétence.

En règle générale, dans les sociétés fragmentées, les engagements constitutifs des rapports entre la presse et les publics naissent moins de l'observation de principes généraux, de la soumission à des critères professionnels collectifs ou d'accords fondamentaux sur des valeurs communes

globales (i.e. sociétales), mais eu égard au degré de communion entre les certitudes idéologico-partisanes des émetteurs et celles des récepteurs. Un fond commun de croyances et une communauté d'esprits, auxquels s'identifient des classes sociales ou des familles culturelles et politiques particulières, asseoient et renforcent la confiance que les lecteurs ont dans les journalistes ou même la compétence qu'il leur reconnaisse. Paradoxalement, si cette communion de pensées prédispose les rédacteurs à respecter des conformismes politico-idéologiques, ceux des journalistes voulant se distinguer s'évertuent de bon coeur à mettre en vitrine des qualités personnelles: le brillant du style et l'élégance de la rhétorique empreints de subjectivisme suppléantent la découverte des "faits" ou la soumission à des régles standards de travail journalistique caractéristiques des écritures de presse à prétention universaliste. *A contrario*, des représentants du "nouveau journalisme" américain des années soixante illustrent cette quête pour s'écarter des contraintes d'un journalisme très professionnalisé. Ceux-ci, en revendiquant bien haut le droit au subjectivisme et celui d'exprimer les sentiments de communautés particulières entendaient rompre avec l'objectivisme factuel associé au professionnalisme dominant.[40]

Par suite, dans les sociétés idéologiquement et politiquement fragmentées, nombre des indicateurs repères d'un rapport d'influence à vocation universaliste perdent de leur relief.[41] presse, réifiée comme un acteur collectif et épinglée dans des formules majuscules qui oblitérent les dénotations individuelles ("The Press", "The Media"), est une institution sociale moins saillante sur la scène publique.[42] Ainsi, les spécialistes, qui auscultent la confiance des citoyens américains dans les institutions fondamentales de l'Amérique, encartent-ils depuis des décennies les objets "Press" ou "Media" dans des baromètres d'opinion ,tandis que cet usage est plutôt rarissime en Europe; du reste, sur le vieux continent, les opinions publiques accordent moins d'importance à la Presse comme institution. De même, les attributs, les sentiments et les organismes qui "collectivisent" l'identité sociale de la profession sont plus faibles. Avec une ampleur peu commune dans des pays comme la France ou l'Italie, des cérémonies rituelles de distribution des prix et accessits louent en Amérique les performances des journalistes et des mass-media; sur ce continent le milieu de la Presse embellit volontiers des épisodes (les dossiers du Pentagone, Watergate) pour les convertir en événements qui deviendront des mémoriaux d'histoire collective; la profession, dans des portraits – standards ou exemplaires, popularise les modèles du grand journalisme américan; de même un réseau d'écoles et d'instituts de formation, dont l'ampleur est sans doute unique sur la planète, met en forme et diffuse les règles de l'esprit du travail de presse. De fait bien que les journalistes américains n'évaluent pas d'une façon similaire leur métier, n'empêche, le plus grand nombre s'identifient à leur profession et se reconnaissent dans ses principes fondamentaux; à l'opposé, les collègues italiens ou français font

d'abord allégeance aux entreprises de presse qui, plus que la profession, constituent leur premier centre d'identité. En plus de cela, dans un contexte d'influence particulariste, la presse se montre moins soucieuse de limiter, de mettre à l'écart ou d'identifier comme tels les "papiers" de commentaires ou d'opinions; du reste, à la différence de la spécialisation anglo-saxonne, les services et les rédacteurs assurent fréquemment des rôles simultanés de "chroniqueurs" et de "reporters".

Lorsque, sous l'invocation de l'autorité du professionnalisme, le journalisme qui se prévaut d'un rapport d'influence universaliste aspire à des états *d'autonomie* des mass-media vis-à-vis des milieux environnants, les organes de presse à vocation particulariste s'inscrivent dans des systèmes différents de relations mais qui *à bien des égards ne sont pas moins légitimes*. Ces remarques s'appliquent tout particulièrement aux mécanismes d'institutionnalisation des liens avec la societé politique.

Dans l'état d'incertitude qui caractérise ces derniers, les partenaires, qu'il s'agisse des reporters ou des élites, n'ignorent rien des panoplies d'actions susceptibles de stabiliser les relations. D'un côté, les membres de la société politique disposent à l'adresse des mass-media d'un arsenal de moyens pour formaliser les rapports. D'une part, les élites peuvent recourir à des stratégies visant à *augmenter la dépendance* des moyens d'information: (i) usage de la contrainte sous la forme de mesures réglementaires ou législatives; (ii) manipulation de ressources matérielles ou d'informations utiles au fonctionnement des firmes ou d'intérêt pour les journalistes et susceptibles de contreparties. Mais, d'autre part, à l'inverse ou d'une manière complémentaire, la société politique peut désirer *accroître son indépendance* vis à vis des organes de presse. Quoique les mass-media soient des canaux privilégiés d'accès vers les citoyens, les élites jouissent, le cas échéant, de (iii) moyens alternatifs de contacts avec les publics grâce aux ressources des organisations de masse (partis, syndicats, associations) ou aux facilités offertes par les techniques modernistes ou traditionelles de propagande (affichage, mailing, banques téléphoniques); les acteurs politiques savent aussi jouer avec des (iiii) alternatives d'action quand ils mettent en concurrence les reporters et les firmes; enfin les élites peuvent s'engager dans des (iiiii) manoeuvres de dénigrement, c'est-à-dire faire usage, de temps à autre des éléments d'autorité associés à leurs fonctions pour miner les rapports d'influence des mass-media avec les publics.

Sur l'autre versant, les ressources à la disposition de la presse ne sont pas moins diverses. Les mass media cherchent à faire croître la dépendance des élites ou à l'inverse ils tentent d'augmenter leur indépendance: (i) les entreprises se saisissent de textes constitutionnels législatifs ou de décisions judiciaires pour neutraliser ou contraindre la gente politique; (ii) la presse excelle, ici et là, dans le jeu des postures d'indifférence: les reporters excluent des informations du champ des nouvelles, ou bien ils les relèguent à l'arrière-plan, (iii) de même le monde journalistique n'ignore pas les

manoeuvres des alternatives d'action: les rédacteurs chevronnés placent les sources de nouvelles dans des situations de concurrence; les journalistes fabriquent – parfois de toutes pièces – des scènes d'interpellation des élites par la publication de données inédites obligeant, tout du moins l'espérent-ils, les acteurs politiques à répondre.

Néammoins, dans un tel contexte d'interdépendance stratégique – c'est à dire que les activités et les décisions des uns et des autres dépendent des conduites réciproques – les modes d'institutionnalisation des rapports varient. Ainsi, les situations d'influence particulariste sont-elles favorables à des stratégies soulignant la dépendance des mass media vis à vis de la société politique, par suite d'un usage étendu des liens de contrôle et de la proximité idéologico-partisane des organes de presse avec des secteurs de la société politique. En revanche, prendre des distances vis à vis des cercles politiques est plus délicat pour les mass-media puisque il semble plus difficile qu'ils se prévalent de l'autorité d'une "professionnalisation" bien constituée et riche d'un soutien consensuel. Faute d'accords généralisés sur des règles, sur des standards permettant d'evaluer d'une manière raisonnable les performances, le professionnalisme particulariste des uns cohabite ou s'oppose au professionnalisme particulariste des autres. Du reste, le législateur, conscient, à l'occasion, de ces états de fait, peut prendre des mesures qui institutionnalisent un statut particulariste de la profession journalistique (par exemple, la clause de conscience en France, et en Italie).

Toutefois, quoiqu'en disent des publicistes porte parole d'un certain journalisme anglo-saxon, dans les pays où les rapports d'influence particulariste sont monnaie courante, le voisinage des mass-media avec des puissances politiques ne sape pas *ipso-facto* la légitimité des relations lorsque des journalistes ou des entreprises se réclament de solidarités partisanes ou collectives – comme l'Etat-nation. "L'*Humanité*, écrit son ancien rédacteur en chef, "est l'organe central du parti, c'est-à-dire qu'il a un rôle d'éducation, d'organisation, de liaison"[43]; à bien y voir, G.Pompidou disait à peu près la même chose quand l'homme d'Etat identifiait la radio-télévision publique à "la voix de la France".[44] Par suite, dans ces circonstances, les perturbations du fonctionnement normal des marchés ou l'encadrement de l'activité des mass-media n'affichent pas nécessairement une nature discrétionnaire exorbitante, de même, lorsque les journalistes se conforment aux injonctions d'acteurs politiques qui les surveillent et les sanctionnent ces pratiques peuvent laisser indifférentes les opinions publiques ou ne pas les choquer outre-mesure. Du reste, assurer les conditions d'existence d'un pluralisme de rapports d'influence particularistes tient au coeur de certaines sociétés nationales-étatiques qui y voient une valeur essentielle-ainsi en est-il de la Suède qui soutient et désire maintenir la diversité politico-idéologique de la presse écrite ou des pays organisant l'accés du journalisme engagé sur les ondes[45]. Ou bien encore: des collectivités, traversées de désaccords fondamentaux y compris parfois de "dissen-

sus" rejetant les bases de l'ordre social dominant ou anticipant des crises susceptibles de mettre en cause la sécurité nationale, reconnaissent comme légitimes la subordination des mass-media aux directives des autorités publiques.

Si les situations d'influence particulariste réduisent les données d'incertitude qui se glissent dans les relations mass-media/société politique, ces éléments occupent un rôle de premier plan dans les contextes d'influence universaliste. Autrement dit, les phénomènes d'interdépendance stratégique possédent beaucoup plus d'ampleur. Cette situation caractérise par exemple une campagne électorale présidentielle américaine. Tout candidat entend se faire connaître, mobiliser des citoyens et à terme: vaincre. Bien que l'Etat-fédéral mette à la disposition des protagonistes des ressources financières facilitant l'usage des techniques du marketing politique, rendu d'autant plus nécessaire par la faiblesse des appareils partisans, l'accés aux ondes reste du domaine des grandes chaînes de radio-télévision privées soucieuses d'autonomie.

Ces circonstances facilitent la mise en place d'un système d'échange social qui à grands traits peut être réduits à deux flux principaux. Dans l'esprit des candidats: nourrir les journalistes d'informations dans des conditions optimales de réception; en sens contraire, du côté des reporters: "avoir accès", selon la formule consacrée, aux candidats et à leur entourage. Or, un banal transfert de ressources ne peut seul activer le système d'échange par suite d'une part de l'incertitude qui s'y glisse, d'autre part, si les partenaires du dit système s'adressent aux mêmes publics, les rapports qu'ils entretiennent avec ces auditoires ne coincident pas toujours. Les protagonistes n'entrent dans le jeu que si l'interaction se déroule sous des aspects convenables et acceptables; les partenaires acceptent les concours réciproques sous la réserve que les rôles sont exécutés selon des principes de rectitude comme ceux attachés à la "professionalisation", assurant d'une part la confiance des élites et d'autre part garantissant l'autonomie professionnelle des mass-media. De tels contextes incitent les élites politiques à reconnaître dans la presse un *acteur-partenaire* au lieu d'y voir un banal *agent de transmission* à mettre et à garder sous contrôle.

A ces cas de figure des rapports presse-société politico-étatique vont correspondre des types de politisation de l'industrie des media. De fait si, d'une manière, ou, d'une autre, les moyens d'information s'insérent *de facto* dans le jeu politique, il faut néammoins distinguer cette situation des rôles dans lesquels les mass-media sont des acteurs explicites de la lutte politique. Dans un contexte d'influence particulariste, la politisation des entreprises ou des journalistes peut être dite *sujette*: ceux-ci s'engagent légitimement et au vu de tous dans les luttes politiques puisqu'ils placent leur ouvrage sous l'autorité de principes ou d'acteurs idéologico-partisans, mais aussi sous celle de communautés spécifiques. Au contraire, dans les configurations d'influence universaliste, la politisation se veut *autonome*:

le professionnalisme dont ils se flattent, les mandats collectifs dont ils se réclament, autorisent les mass media à entrer, le cas échéant, dans les jeux mêmes de la société politique sans pour cela se confondre *ipso-facto* avec ses aspects partisans les plus marqués.

Enfin, quelles formes d'institutionnalisation vont revêtir les liens avec le secteur économique? Le contrat, outre des fonctions économiques comme celles de son rôle dans la formation des prix, constitue une procédure de contrôle permettant de prévoir, parfois avec un grand luxe de détails, les conduites attendues; mais il faut aussi y voir, quoique les dispositions puissent être simplement tacites, un cadre normatif délimitant les objets légitimes des clauses contractuelles. L'influence à dominante universaliste n'exclut pas, tant s'en faut, l'usage des ressources du secteur économique mais sous la réserve que celui-ci n'interfère pas outre-mesure dans la confection des nouvelles, sinon quelques-uns des engagements fondamentaux risqueraient de s'effriter. En effet, ces firmes de presse espèrent que la distance vis-à-vis du secteur économique et la présence de clauses limitées aux transactions marchandes leur permettent de mieux ancrer les rapports d'influence universaliste auprès des publics; ces attitudes sont aussi susceptibles de préserver les entreprises des interventions éventuelles des élites politiques. En effet ces dernières tirent parfois avantage de l'emprise réelle ou exagérée des milieux économiques et financiers pour prendre des mesures correctrices ou pour critiquer sur la place publique le fonctionnement "réel" des mass-media. Dans ces circonstances, les contextes d'influence universaliste prédisposent les firmes à promouvoir des règles ou à installer des modes d'organisation qui rendent visible la séparation des univers marchands et des produits journalistiques.

D'ailleurs, ce rigorisme présent au *Monde* ou chez ses confrères américains n'effraie pas nécessairement les annonceurs, puisque, au nom d'un intérêt bien compris, il leur importe souvent que les "supports" jouissent d'une bonne confiance auprès des publics. Cette réserve rend donc *des* mass-media sensibles aux risques de dépendance vis-à-vis des milieux extérieurs, et de fait des journaux ou des stations de radio-télévision se montrent soucieux par exemple de diversifier les budgets publicitaires. De même, l'influence universaliste valorise les critères de bonne gestion dans la mesure où ceux-ci sont perçus comme permettant de contenir les dépendances vis-à-vis du secteur économique ou d'éloigner la tentation de s'en remettre aux bailleurs de fonds de la société politico-étatique. Enfin cette volonté d'autonomie invite les firmes à regarder de près et d'une manière active des stratégies d'audiences comme celles d'atteindre des publics de prestige propices à consolider le rapport d'influence universaliste.

Sans doute ces considérations n'échappent-elles pas aux firmes de presse qui privilégient les rapports d'influence particulariste. Toutefois, ces pratiques risquent, d'une part d'être moins fréquentes ou pregnantes,

d'autre part les clauses des contrats n'excluent pas des dispositions – lesquelles ne nécessitent pas d'être consignées par écrit – étrangères à l'univers marchand. Des financiers, des chevaliers d'industrie, des serviteurs de l'Etat ou des responsables d'organisations politiques accordent des ressources à certaines firmes de presse, par suite précisément des modalités particularistes de leur rapport d'influence. Si, à l'évidence, les sociétés fragmentées d'un point de vue idéologique et politique facilitent de telles pratiques, parfois, la nature même des rapports particularistes les renforce: les publics atteints ne possèdent pas des dimensions suffisantes pour assurer des revenus adéquats, par suite des frontières politico-idéologiques qui les séparent. L'assistance externe des bailleurs de fonds devient tout naturellement nécessaire, elle peut sembler aussi légitime puisqu'une telle bienfaisance assure la survie du particularisme des entreprises que l'on valorise.

Dans ces conditions, les critères d'efficacité économique semblent moins impératifs, les entreprises peuvent se satisfaire de scores de performances plus lâches puisque d'autres considérants s'infiltrent à l'occasion dans les conceptions de la bonne gestion. Si les rapports d'influence universaliste prédisposent à des stratégies en quête de marchés rentables sinon étendus, par contre, les entreprises "particularistes" regardent plus volontiers les parts de marchés définies en termes d'accords ou de proximité politico-idéologiques. A contrario, lorsque ces firmes aspirent ou décident de prendre des distances vis-à-vis des rapports d'influence particulariste, elles risquent de faire place à des crises d'identité ou de survie nécessitant du savoir faire managérial et rédactionnel.

## Propriétés systémiques de l'entreprise de presse

Ce schéma perçoit donc, dans *l'interdépendance* des mass-media avec les milieux qui les environnent, une caractéristique non seulement essentielle mais aussi complexe. Regardons sous cet angle une entreprise placée dans un contexte d'influence universaliste. Avec plus ou moins de bonheur, elle s'efforce à ce que ses produits journalistiques ne soient pas identifiés par les publics avec l'un ou l'autre des bords partisans; dans ce cadre les manoeuvres de "professionalisation" constituent des pièces essentielles pour offrir des produits susceptibles de plaire à un volume optimal de lecteurs ou d'auditeurs, pour le moins qui ne provoquent pas d'ostracisme à leur égard. Ce faisant, les stratégies commerciales et de production journalistique s'épaulent: les attitudes et les conduites associées à l'influence de type universaliste soutiennent des démarches gestionnaires visant à optimiser les revenus – à l'inverse les stratégies commerciales appellent de leurs voeux des produits journalistiques exempts d'un particularisme outrancier. En cas de succès, ces activités créent des flux de

revenus favorables à l'amélioration des produits, les performances rejaillissent sur les tarifs publicitaires, et de tels résultats peuvent permettre aux firmes de garder leurs distances vis-à-vis du milieu économique ou de la sphère politique.

Mais ces flux ne se contentent pas de créer simplement des ressources utilitaires puisque simultanément leur fonctionnement peut activer ou produire des ressources *normatives* favorables au bien être économique des entreprises. Lorsqu'une firme valorisant l'influence universaliste rencontre le succès, ce type de rapport avec les publics développe chez les clientèles des sentiments de soutien et de loyauté à son égard. Ces éléments sont des ressources susceptibles de renforcer l'ambition d'autonomie des firmes à l'égard des éventuels débordements des milieux politiques ou marchands. Un autre cas de figure est le suivant: les ressources peuvent être manipulées par des acteurs internes aux entreprises poursuivant des objectifs différents. Ainsi, les manoeuvres de "professionnalisation" ne se limitent-elles pas au champ des rapports avec les publics. Outre-Atlantique, l'audience progressive des phénomènes de "professionnalisation" depuis la fin du 19ème siècle s'accompagne d'une dynamique de développement qui les consolide. Les reporters se saisissent des valeurs professionnelles pour asseoir l'identité sociale du groupe des journalistes, mais aussi pour s'émanciper de la toute puissance des *publishers*; de façon réciproque, propriétaires et dirigeants invoquent le "professionnalisme" pour contenir les emballements et les intérêts des salles de rédaction. Lorsque, aux Etats-Unis, l'économie de marché florissante des années vingt découvre les techniques de la publicité moderne et des relations publiques, quand depuis la Première guerre les gouvernements perfectionnent les moyens de propagande, patrons de presse et journalistes réagissent par de nouvelles manifestations de professionnalisme dont les plus ostensibles sont assurément la standardisation de l'écriture de presse et la parution de codes de déontologie.

Ensuite, si ces phénomènes d'interdépendance s'accompagnent d'une part de *transferts* de ressources, d'autre part ils correspondent à des processus de *transformation*. La circulation des ressources permet de fabriquer des "produits" mais les mécanismes de transformation "d'inputs" en "outputs" ne manifestent pas toujours un caractère proprement technique. Les recettes publicitaires permettent de publier ou d'émettre, mais à la condition expresse, dans le cas d'un rapport à dominante universaliste avec les publics, que les transactions avec le secteur économique soient correctes eu égard à des critères de rectitude pour la plupart étrangers à la sphère marchande, sinon elles risquent d'être perçues comme des cas de corruption. Précisément, il y a corruption, comme le voit bien un auteur parsonien N. Smelser, lorsque le "crossing over" des ressources échangées est perçu comme "repréhensible" au regard d'un cadre normatif.[46] La presse acquiert, mobilise des ressources financières et elle les convertit grâce à la

médiation d'éléments d'une nature différente — ici, d'une qualité normative — qui peuvent être originaires d'autres milieux. L'interdépendance n'est donc pas une simple affaire de coordination technico-fonctionnelle, elle requiert aussi la *compatibilité* normative des ressources et des rapports. Ainsi l'autonomie revendiquée vis à vis de la classe politique se dévalorise-t-elle si les mass-media s'en réclamant se conforment à l'excès aux injonctions des puissances financières ou s'ils suivent de trop près les desiderata des publics. L'importance de ces phénomènes se montre à découvert avec les monopoles d'Etat dans lesquels la compatibilité ne va pas de soi. D'ordinaire, les monopoles de radio-télévision prétendent répondre, au nom de l'intérêt général, aux attentes collectives des citoyens, mais, précisément, le statut de monopole, sauf exceptions comme la B.B.C., les place sous l'emprise des mécanismes de contrôle de la société politico-étatique. Cette ambiguité fondamentale impose *de facto* un caractère problématique aux rapports d'influence susceptibles de s'établir. Dès lors, de telles entreprises aiment exhiber avec plus ou moins de succès des attitudes de neutralité qui, d'une manière factice ou réelle tentent d'oblitérer les traits voyants des liens de contrôle. Les cas de monopoles de fait, comme ceux en France de quelques quotidiens régionaux illustrent d'autres démarches stratégiques. Ces organes de presse, placés dans une situation relativement hégémonique entendent atteindre tous les publics, bien que ceux-ci puissent être idéologiquement fragmentés. Dés lors, faute des avantages d'une société consensuelle et d'une autorité professionnelle suffisante, ces publications s'emmitouflent volontiers dans un apolitisme ostensible, elles se montrent respectueuses des pouvoirs, singulièrement lorsque les activités de ces derniers sont susceptibles de prêter aux controverses; ce faisant, de tels organes de presse réduisent la nature problématique des rapports d'influence avec les publics associée aux positions de monopole de fait. Plus généralement la "neutralisation du produit" facilite "l'indépendance relative des opinions politiques des lecteurs par rapport aux prises de position politiques du journal".[47]

Ces analyses montrent que les exigences de compatibilité et de cohérence dépendent de *points-clés* comme celui du type de rapports entretenus avec les publics. Dans ce cadre d'analyse, les relations-types avec les lecteurs sont des éléments essentiels de la légitimité *de principe* des liens entretenus par les organes de presse avec les milieux politiques ou économiques sans pour cela être assurée de les déterminer en pratique. De fait, toutes ces propriétés d'interdépendance et de transformation, de compatibilité ou de cohérence du système des rapports constitutifs de l'entreprise de presse, présupposent l'action de phénomènes de contrôle nécessaires à une intégration "systémique". A maintes reprises, les analyses précédentes révèlent des jeux de contrôle, lorsque par exemple l'usage des ressources économiques nécessitent l'intervention d'éléments normatifs. Si la mise à jour des contrôles peut faire, non sans difficultés, l'objet d'un constat

empirique, la question essentielle est bien sûr celle de mettre au clair leurs modalités de fonctionnement et leur importance respective. L'entreprise est délicate par suite, d'une part, de la variété des situations, et par suite, d'autre part, des fluctuations au cours du temps. La littérature d'inspiration managériale suggère néanmoins une voie d'enquête pour simplifier ce problème: dans quelle mesure les liens avec les environnements contribuent-ils, comme disent les experts en gestion, à la "survie" des firmes?[48] Selon cette perspective (i) les rapports-types remplissent des fonctions de contrôle les uns vis-à-vis des autres; (ii) l'intensité des régulations dépend de la contribution des rapports-types à la "survie" *perçue* des entreprises.

Ce critère de contribution[49] fluctue eu égard à:

(i) *l'importance* donnée par les dirigeants des entreprises aux buts qu'ils entendent atteindre. Certaines valorisent des performances économiques lesquelles du reste varient: rentabilité, croissance, conquête de positions dominantes; d'autres ou les mêmes sont soucieux d'offrir un journalisme de prestige ou de participer à des luttes politico-idéologiques;

(ii) à la *centralité* des rapports dans le fonctionnement des firmes. Bien qu'une entreprise considère comme primordiale un rapport d'influence universaliste avec ses publics, ce dernier n'efface pas, tant s'en faut, les contraintes qui peuvent être décisives pour une bonne gestion financière ou technique.

(iii) au *degré de substitution* ressources et entre les rapports. Par suite de succès auprès des audiences une entreprise peut être tentée de prendre ses distances vis à vis des bailleurs de fonds ou des surveillants politiques; de même, des mass-media ne réussissant pas à générer des ressources financières suffisantes recherchent des "mécènes" qui en retour imposent des conditions susceptibles d'infléchir leur travail journalistique.

Cette stratégie d'enquête, appelant des études empiriques et une vision globale du fonctionnement des firmes[50], dispense de postuler une hiérarchie stable et univoque des régulations, puisque, d'une part, elle repose sur une variété de ressources, et que, d'autre part, des contingences risquent sans cesse de la perturber. Parfois, les ressources tirées de la vente s'avèrent insuffisantes et pour, faire face à la croissance des coûts de production des entreprises se trouvent dans l'obligation temporaire d'accroître leur dépendance vis-à-vis de certains budgets publicitaires. L'instabilité des régulations dans une entreprise risque d'être d'autant plus forte lorsque les fonctions de contrôle correspondent à des enjeux de parties de pouvoir disputées par quelques-uns de ses partenaires — fût-ce, le cas échéant avec le soutien d'acteurs sociaux externes aux firmes. Au sein d'un journal ou d'une station de radio, des catégories de dirigeants ou d'employés jouissent d'une certaine autonomie d'action qui entre en conflits avec les volontés d'autres services; des groupes comme celui des journalistes prétendent exercer une influence déterminante sur les conduites des entreprises -serait-ce au prix d'antagonismes sévéres avec les "commer-

ciaux" ou avec les "financiers". Dans ces conditions, les entreprises de media recherchent tout d'abord des *équilibres internes* qui, par suite du mélange de rationalités les traversant, sont plus délicats à atteindre que ceux des firmes typiquement industrielles ou commerciales. Du reste, cette quête se traduit par des procédures variables d'organisation et de direction: jusqu'à une période récente le journal *le Monde* se considérait comme un "tout", dont le fonctionnement se voulait subordonné à la suprématie d'une certaine idée du journalisme de prestige; par suite dans son mode quotidien d'organisation, les rôles des différentes fonctions s'entremêlaient. De son côté, le *Washington Post*, avec une ambition comparable, segmente les sphères d'action: quelle que soit l'interdépendance des secteurs techniques, commerciaux et de la *newsroom*, le journal américain favorise leur autonomie réciproque, propice, pense-t-on, à l'épanouissement des compétences respectives.

En second lieu, les entreprises de media recherchent non seulement des équilibres internes, elles tentent aussi d'atteindre des positions dominantes. Cette tendance se traduit en particulier par des efforts pour atteindre des états d'oligopole ou de monopole de fait permettant de mieux maîtriser des ressources essentielles de fonctionnement.

Bien qu'il s'agisse en apparence d'un paradoxe, les contextes d'influence universaliste favorisent de telles tendances. L'exigence d'autonomie dont se réclament les mass media les rend sensibles aux scores des performances économiques et financières; dès lors, la surveillance des coûts de production, la croissance du capital fixe, incitent les entreprises à rechercher des économies d'échelle ou à trouver des politiques d'intégration et de diversification; dans le même sens, les stratégies commerciales soucieuses de conquérir et de garder des "clientèles", s'efforcent de rendre ces dernières un tant soit peu "captives", puisque les publics risquent toujours de faire défection. A l'évidence, ces phénomènes se retrouvent dans les contextes riches en rapports d'influence particularistes; toutefois, leur ampleur risque d'être plus faible; mais surtout, les mouvements monopolistiques ou oligopolitiques ne semblent pas résulter du simple jeu de la conjonction de l'autonomie journalistique et du souci des performances économiques et financières. Plutôt, les évolutions de tendances vers des états de monopole ou d'oligopole profitent du concours voire même de l'intervention plus ou moins discrète d'acteurs de la société politique ou du secteur économique favorables à ce que des organes de presse puissent s'installer dans des positions dominantes. Par suite, l'entrée en lice de la puissance publique ne repose pas sur des légitimités identiques bien qu'elles soient formellement similaires. Dans un contexte d'influence universaliste le législateur intervient au nom de principes généraux pour faire respecter des règles du jeu comme celle de la concurrence des idées et des intérêts; lorsque la société politique et le secteur économique sont soucieux de contrôler les medias, l'usage des mêmes

principes risquent d'apparaître sous le visage de manoeuvres en quête d'ojectifs particuliers.[51]

## Conclusion

Quel est l'intérêt de telles démarches théoriques pour le champ d'étude des mass media? Dans le cas présent, la mise à jour de dimensions et de propriétés caractéristiques des entreprises de presse invitent à *d'authentiques* études comparatives.[52] Ainsi l'analyse théorique permet-elle, nous semble-t-il, de contenir les facilités des explications en termes de "caractères nationaux"; *les singularités qui se dégagent des mises en rapport comparatifs résultent moins d'idiosyncrasies nationales que de la combinaison singulière de propriétés ou de dimensions générales et communes autorisant les mises en parallèle.* En vérité, du moins est-ce notre voeu, à la lumière du cadre d'analyse développé dans ces pages, les journalismes du continent européen ne semblent pas des genres déviants; par contre, *un tel schéma fait plutôt ressortir l'extraordinaire singularité du journalisme nord américain lorsqu'il revendique une nature universaliste.*[53] Ce travail suggère de nouvelles perspectives d'enquête sur la *rationalité* des mass-media considérés comme des organisations. L'étude s'ouvre sur une vision utilitariste des organes de presse qui se soumettent à une rationalité technico-économique permettant d'établir des rapports entre les avantages et les inconvénients des actions; le mode d'opération inspirant les conduites de la rationalité technico-économique correspond au calcul. Certes, et à n'en pas douter; toutefois, l'enquête montre la nature incomplète, pour ne pas dire fallacieuse, de cette façon de voir. Simultanément les mass media, selon des modalités variables, se veulent *corrects* – pour remettre en mémoire une formule utilisée par H. Simon dans sa jeunesse – au regard de valeurs, de principes ou de pratiques.[54] En d'autres termes le langage de l'efficacité s'interpénètre avec une rationalité normative qui oblige les mass-media à être *conformes* à des règles, à des conventions ou à des normes définissant les contours *légitimes* du travail journalistique et des activités des firmes.[55] W. Lippmann suggère fort bien cette nature quand il écrit:

> La presse n'est pas une affaire commerciale banale, d'une part ses produits sont en général vendus au-dessous des prix de revient, mais surtout la société plaque sur elle des standards de bonne conduite qui n'ont pas d'équivalents dans les autres secteurs industriels. A bien y regarder tout se passe comme si un journal devait être assimilé à une église ou à une école.[56]

Dès lors, un organe de presse sera dit *rationnel*, non seulement lorsqu'il tente d'optimiser l'usage de ressources, mais aussi quand celui-ci s'efforce d'être simultanément conforme à des critères de rectitude soumis à des procédures distinctes de régulation sociale: l'adhésion, la loyauté à des principes, ou bien encore, à des sentiments d'obligation.

## Notes

* Ce texte reprend pour l'essentiel une communication présentée à l'Ecole d'été de l'Institut Universitaire Européen de Florence (8-26 juillet 1985) et préparée dans le cadre du projet "Analyse comparative des politiques de communication". Nous tenons à remercier pour leur aide précieuse l'I.U.E., les participants de l'Ecole d'été, et pour leurs commentaires et collaboration, M.A. Catotti, F. Bourricaud, Ph. Schlesinger.

[1] Pour des raisons de clarté et de style nous ferons usage, outre l'expression "entreprise de presse", des termes "media", "mass media", "moyens d'information", "organes de presse". De même nous emploierons indifféremment les mots "reporters", "journalistes", "rédacteurs".

[2] Cf. par exemple L. Engwall, *Newspapers as Organizations*, Farnborough, Hants, Saxon House, 1978; D. McQuail, *Mass Communication Theory*, Londres, Sage Publications, 1983.

[3] Cf. par exemple E. Yuchtman, S.E. Seashore, "A system resource approach to organizational effectiveness", *American Sociological Review*, 32, pb 891-903.

[4] Dans ce texte, milieux et environnements sont synonymes.

[5] Cf. par exemple, M.A. Dubick, The Organizational Structure of Newspaper, *Administrative Science Quaterly*, 28, 1978-79, pp.

[6] S. Hess, *The Washington Reporters*, Washington D.C., The Brookings Institution, 1981; Leon V. Sigal, *Reporters and Officials*, Lexington, Mass, D.C. Heath, 1973.

[7] Cf. les références citées par Bruce M. Owen, *Economics and Freedom of Expression*, Cambridge, Ballinger, Publishing Company, 1975.

[8] Précisons par quelques remarques l'ambition et les limites de ce travail. Le schéma d'analyse offert dans les pages suivantes ne prétend pas rendre compte en détails du fonctionnement concret des entreprises de presse. Il s'agit d'un *modèle* laissant à l'arrière-plan les corpus d'observations et d'analyses qui ont servi de base à son élaboration. Ce cadre de réflexion sous-tend par exemple notre étude comparative du *Monde* et du *Washington Post* (Paris, P.U.F. 1985).
D'autre part, l'étude porte principalement sur l'une des activités des entreprises de presse: *la production de nouvelles*. L'analyse est centrée sur la *presse écrite* car les media audio-visuels nécessitent, nous semble-t-il, des réflexions complémentaires.

[9] En particulier nous empruntons à l'article de T.Parsons paru dans la première livraison d'*Administrative Science Quaterly*, "Suggestions for a Sociological Approach to the Theory of Organizations", 1957, 1, 1, pp. 63-85. Sur Parsons, cf. en particulier, F.Bourricaud, *L'Individualisme Institutionnel*, Paris, P.U.F., 1977.

[10] Signalons néammoins les études d'O. Burgelin, *La Communication de masse*, Paris, S.G.P.P. 1970, de H.M.Johnson, "The Mass Media, Ideology and Community Standards, in J.Loubser, *Explorations in General Theory*, New York, Free Press, 1976, pp.609-638.

[11] P.M. Hirsch suggère une approche de ce type dans "Occupational, Organizational, and Institutional Models in Mass Media Research", in *Strategies for Communication Research*, London, Sage Publications, 1977, pp. 13-42.

[12] Cf. à ce propos les remarques de R. Escarpit, *Théorie générale de l'information et de la communication*, Paris, Hachette, 1976; M. Woolf, *Teorie delle comunicazioni di masse*, Milan, Bompiani, 1985.

[13] *Op.cit.*

[14] *Op. cit.*

[15] *Id. op.cit.*, p.31.

[16] Cf. par exemple U. Eco, *A Theory of Semiotics*, Bloomington, Indiana University Press, 1946; M. Gottediener, "Hegemony and Mass Culture: A Semiotic Approach", *American*

*Journal of Sociology*, 1985, 90, 5, pp. 979 – 1001; L. Prieto, *Etudes de linguistique et de sémiologie générales*, Génève, Librairie Droz, 1975.

[17] E.C. Hughes (sous la direction de), *Society. Collected Papers of Robert E. Park*, Glencoe, Free Press, 1955.

[18] *De la Démocratie en Amérique*, Paris, Garnier-Flammarion, 1981, tome II, p.146.

[19] P. Blau, *Exchange and Power in Social Life*, New York, J.Wiley, 1967.

[20] J. Morgan, *La Presse et l'Etat: la réglementation de la presse écrite dans douze pays occidentaux*, Québec, Ministere des Communications du Québec, 1981.

[21] *Id.*

[22] *op.cit.*

[23] Cf. à titre d'illustration du modèle de l'échange social, M.B. Grossman & *al*, "The Media and the Presidency: an Exchange Analysis", *Political Science Quaterly*, 1976, 91, pp.435-470.

[24] A.M. Okun, *Prices and Quantities; a macroeconomic analysis*, Oxford, Basil Blackwell, 1981.

[25] C'est-à-dire selon l'expression d'O.Burgelin "la communication qui est destinée à faire pression", *Op. cit* p.262.

[26] Nous nous inspirons pour ces distinctions de l'analyse de Parsons sur l'influence. "On the Concept of Influence" in *Politics and Social Structure*, New York, The Free Press, 1969, pp.352-404.

[27] Cf. l'inventaire des travaux recensés par N. Lin, *The Study of Human Communication*, The Bobbs – Merrill company, Inc. 1971, Indianapolis.

[28] P. Albert *La Presse française*, Paris, La documentation française, 1983, p.28.

[29] H. Becker, "The Nature of a Profession", *Sociological Work*, Londres, Penguin, 1971, p.92.

[30] T.J. Johnson, *Professions and Power*, Londres, Mac Millan, 1972.

[31] Cf. à ce propos, T.J. Johnson, *Op. cit.*

[32] Des travaux récents permettent d'éclairer ces développements socio-historiques dans l'exemple de la presse américaine J.C. Alexander, "The Mass News Media in Systemic, Historical and Comparative Perspective", in E. Katz, *et al*, *Mass Media and Social Change*, Londres, Sage, 1981, pp. 17-51; D. Shiller, *Objectivity and the News*, Philadelphia, The University of Pennsylvania Press, 1981, M. Schudson, *Discovering the News*, New York, Basic Books, 1978; G. Tuchman, *Making News*, New York, Free Press, 1978.

[33] Cf. A. Abbott, "Professional Ethics", *American Journal of Sociology*, 1983, 88, pp. 855-885.

[34] T. Burns, *The BBC, Public Institution and Private World*, Londres, The Mac Millan Press, 1977.

[35] Cf. à ce sujet les remarques sur la notion d'acteur collectif comme construction sociale de M.E. Spencer, "Images of Groups", *Archives Européennes de Sociologie*, XVI, 1975, pp.194-214.

[36] *Op. cit.*, p.5.

[37] Ce phénomène caractéristique des sociétés techniciennes est décrit en termes généraux par J. Habermas, en particulier dans, *Toward a Rational Society*; Boston, Beacon Press, 1970.

[38] S.P. Huntington, *American Politics, The Promise of Disharmony*, Mass. The Belknap Press of Harvard University Press, 1981; ou bien encore dans une perspective différente: T. Gitlin, *The Whole World is Watching*, Berkeley, California University Press, 1980.

[39] L'importance des dimensions des systèmes politiques sur les mass media est soulignée par J.G. Blumler, M. Gurevitch, "Linkages between the Mass Media and Politics", in J. Curran, (sous la direction de) *Mass Communication and Society*, Londres, Edward Arnold, 1977; C. Seymour Ure, *The Political Impact of Mass Media*, Londres, Constable, 1971, G. Grossi, G. Mazzoleni, "Per una interpretazione del rapporto tra Parlamento e sistema

informativo", in *Informazione e Parlamento*, Roma, Camera dei deputati, Ufficio stampa e pubblicazioni, 1984, pp. 135-179.

[40] L'essai de J.C. Merrill rend explicites ces attitudes, *Existential Journalism*, New York, Hastings House, 1977.

[41] Cf. pour plus de détails J.G. Padioleau *op. cit.*

[42] Cf. les resultats d'enquête commentés par R. Rose, "Public Confidence, Popular consent", *Public Opinion*, février-mars 1984, pp. 9-11, 60.

[43] R. Andrieu, *Du bonheur et rien d'autre*, Paris, Stock, 1975, p. 204.

[44] D. Bombardier, *La voix de la France*, Paris, Laffont, 1972.

[45] Cf. S. Hoyer *et al*, *The Politics and Economics of the Press*, Londres, Sage, 1975.

[46] "Stability, Instability and the Analyses of Political Corruption", in B. Barber (sous la direction de) *Stability and Social Change*, Boston, Little Brown, 1971, pp. 7-30.

[47] P. Bourdieu, *La distinction*, Paris, Editions de Minuit, 1979, pp. 516-517.

[48] Cf. par exemple les manuels de H. Mintzberg, *The Structuring of organizations*, Englewood Cliffs, Prentice Hall, 1980; H. Aldrich, *Organizations and Environments*, Englewood Cliffs, N.J. Prentice Hall, 1979.

[49] Ces variables sont adaptées des analyses de D.J. Hickson *et al*, "A Strategic Contingency Theory of Intra-Organizational Power", *Ad. Sc. QuarterlyN, 1971, 16, 3, pp. 219-232.*

[50] C'est dire l'intérêt des études de la "political economy" des mass media entreprises par P. Golding, G. Murdock, *et al.*

[51] Un exemple caractéristique est assurément la législation française sur les entreprises de presse du gouvernement de P. Mauroy.

[52] Nous avons développé ce point de vue dans une autre étude en cours de parution, "L'entreprise de presse comme institution", (1984).

[53] C'est l'un des thèmes majeurs de notre enquête *Le Monde vs The Wasghinton Post, op. cit.*

[54] *Administrative Behavior*, New York, The Free Press, 1945.

[55] J.W. Meyer, et B. Rowan ont attiré l'attention sur ces phénomènes dans "Institutionalized Organizations", *American Journal of Sociology*, 1977, 89, 2, pp. 340-363.

[56] W. Lippmann, *Public Opinion*, New York, The Free Press, 1965, pp. 203-204.

# Democratic Theory and Neo-corporatist Practice

PHILIPPE C. SCHMITTER
Stanford/Florence

By all accounts, democratic theory is in a lamentable state. It has certainly not kept pace with, much less guided, changes which have occurred in the practice of politics over recent decades. It is as if the Western democracies, having defeated their anti-democratic foes in the Second World War (with the help of the Soviet Union) felt no further need to examine critically their changing political practices. Anything these polities did – any public institutions they created, any private arrangements they tolerated – were *eo ipso* democratic since they emerged consensually (or at least without violence or discontinuity) from within polities which otherwise respected basic civic rights and held regular contested elections.

In this context, much of what passed for democratic theory consisted of belated (and often apologetic) efforts at *restatement*, i.e. at justifying after the fact why particular practices or sets of institutions should be considered "democratic" even if their inspiration was purely pragmatic or their intent purely utilitarian. And one must concede that this capacity for *post factum* incorporation has been one of the keys to the endurance of democratic theory. Just think of the institutions which were created for other, occasionally even diametrically opposed, motives that subsequently have become part-and-parcel of our contemporary conception of democracy: parliament, constitutions, mixed government, checks and balances, proportionality, factions or parties, interest organizations, consociational compromises, etc.

Of course, activitists and scholars have dared to criticize some of the more indiscriminate of these "realistic" accomodations on the grounds that they so stretch the concept of democracy that it loses all logical connection with its past meaning. Perhaps the pinnacle of this sort of stretching was reached when the American pluralists extolled the virtues of "apathy" and "slack resources" in explaining why the US polity was both stable *and* democratic. Unfortunately, these critical efforts at *retrieval* often have a reactionary cast to them. They imply (or explicitly urge) that the ensuing (and often well-entrenched) practices be somehow eradicated and the polity be returned to some more desirable (and often unprecedented) state of greater conformity to democratic principles; Since it is usually left

unclear what likely configuration of power would produce such a purge or what possible transformation of citizen attitudes would have to occur for this to be accepted, such exercises remain largely academic. They may stimulate heated discussion among philosophers and ideologues, but few lead to actual experiments in more democratic governance.

This essay is neither an effort at restatement nor at retrieval. It aims at a *reconstruction* of democratic theory. It will attempt to examine critically the implications and consequences of the emergence of neo-corporatist arrangements for the practice of democracy in advanced industrial/capitalist societies without "(indulging) in the fantasy of an unconstrained ideal order".[1] In other words, it accepts the fact that important structural changes have occurred in the way class, sectoral and professional interests are organized and in the way these organizations relate to each other and to the State in polities which have otherwise retained open, competitive electoral processes and civic liberties. However, it neither accepts these changes as democratic because they seem to have been spontaneously created and voluntarily agreed upon, nor does it engage in the wishful fantasy that they represent merely temporary aberrations doomed to collapse in short order before the opposition of democratic forces. Neo-corporatist arrangements are well-entrenched – much more so in some polities than others – and, at least at the level of popular consciousness, they have yet to be rejected as manifestly undemocratic. The task of constructive democratic theory in this context is twofold: (a) to examine critically the impact of neo-corporatist arrangements with an eye to modifying or minimizing them in order to make them more compatible with the enduring principles of democracy; (b) to identify the agents or processes which might possibly promote such a more desirable outcome.

## Democracy

Our obvious first task is to establish as clearly as possible what our standard of democratic performance is. Only then can we convincingly argue that a given arrangement is or is not compatible with it. Democracy is a *principle*, embodied in a set of *rules*, expressed through a complex of *institutions*, which is aimed at establishing or sustaining a certain *quality* of relationship between those who rule and those who are ruled. Its guiding principle is that of *citizenship*, i.e. the right to be treated by fellow human beings as equal and the obligation to respect the legitimacy of choices made by collective deliberation among equals. The *decision rules* which embody this principle have varied a great deal historically – varying from unanimity, to qualified majority, to concurrent minorities, to the presently predominant (but by no means universal) one of simple majority[2] – as have the *eligibility rules* for participation which define who is a citizen and

how he/she can act politically. Secret balloting, universal suffrage, civic freedoms, regular elections and party competition are elements common to virtually all efforts at conceptualizing "modern procedural democracy",[3] while others such as proportional representation, public financing of parties and unrestricted access to information are not (yet) such integral components of the procedural minimum. Needless to say, the *specific institutions*, which have been used to express the citizenship principle and to embody the procedural rules have varied even more over time and across units. Open assemblies, hereditary or appointed chambers, popularly elected presidencies, party caucuses, electoral primaries, worker's councils, advisory committees, federal states, referenda, ombudsmen, – just to name a few – have all been considered legitimate, if different, expressions of democratic institutions at one time or another.

The historically variant procedures and institutions for translating the principle of democratic citizenship into practice are presumably not ends in themselves (or else, why would they have been modified so much and so often?). Rather, they should be regarded as the means for establishing a particular sort of relationship between rulers and rules. The ideal, of course, has always been to abolish this distinction so that citizens could practice self-government, i.e. not only be free and equal in their choices of what course or collective action to take, but free and equal in the carrying out of that course. Once increased size, international threat, decisional complexity, etc. make that ideal unattainable – wherever citizens accept the need for permanent governance by some specialized set of rulers – then the crucial criteria for evaluating democratic performance becomes the *quality* of that ruling relationship. What those criteria should be and in the event of conflicting capabilities which should have priority, have been the subject of much discussion and little agreement. First, it may be useful to distinguish whether the units of reference for evaluating democratic performance are citizens or authorities. In other words, is democracy a quality inherent in the behaviour of the individuals that compose it?, or of the collectivity that governs them? A second evaluative dimension focuses on the process of policy-making and distinguishes between the input and the output aspects of governing. This asks whether a given government is "of the people", or whether it is "for the people" – whether it affords opportunities for citizens to become equally involved, or whether it does things which equally benefit them. Figure I is an effort to depict these contrasting (but not necessarily contradictory) dimensions graphically.

*Participation* is the virtue stressed by those who regard democracy primarily from the point of view of individual citizens playing an active and equal role in the making of collective decisions. Judged from the perspective of public authorities, government of the people should be *accessible* by treating the preferences and demands of all citizens – organized or unorganized, concordant or discordant, precedented or un-

**Figure I**

*The qualities of democratic government*

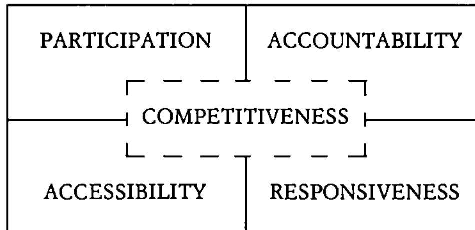
I. The Aspect of Governance

"Of the people" "For the people"

II. The Unit of Reference:

Individual Citizens

Public Authorities




precedented – as equally qualified and worthy of consideration in the making of public choices. If one switches to the second, vertical, axis of what governments do rather than who gets involved, the extent to which individual citizens can hold their rulers *accountable* through some process of regular consultation and non-arbitrary deliberation becomes the most democratic standard one can apply. Finally, *responsiveness* would seem to be the best term for designating the collective property of a democratic ruling class which guarantees that citizen needs will be met and, hence, that the exercise of public power will be legitimated.

Finally and most controversially, *competitiveness* seems to lie at the core of the entire modern democratic process, affecting both individuals and authorities as well as the input and output of governance. Classical democratic theory certainly did not stress this quality. Rather the contrary, the presumption was that free deliberation among equals would produce a singular, ethically superior, consensus ("the public good"). Those who persisted in dissenting from it were expected to leave the community or renounce their citizenship. One quiet revolution in the modern conception of social order – later translated, so-to-speak, into democratic theory – was the notion that communities could persist, even thrive despite stable divergences of belief, interest, lifestyle, etc. in their midst. First, with the tolerance of religious dissent and, later and, however precariously, with respect to status, class, linguistic, ethnic, generation, gender and other cleavages, democracy became virtually identified with the acceptance of social pluralism and, consequently, with the institution-

alization of such cleavages into non-violent competition for office and influence. Presumably, competitiveness stimulates citizens to participate more actively, encourages authorities to be accessible to a greater multiplicity of demands, enforces more accountability on the part of policy-implementors and ensures that collective choices will be responsive to the full range of diversity present in a given society. Without it, the whole matrix of democratic qualities risks degenerating into formalistic ritual, choreographed consent, symbolic distortion and/or generalized apathy.

"Of course!", the reader may protest, "a truly democratic polity should be *all* these things: participatory, accessible, accountable, responsive and competitive. Why bother even to distinguish among them?" One answer is that some of these qualities may be antithetic under particular conditions. Another is that scarcity of resources may force citizens and/or rulers to choose one over the other. But for the purposes of this essay, what is important is that a given new institution or arrangement may contribute to satisfying one of these virtuous qualities – and, therefore, acquire a form of democratic legitimation – while undermining the others. This, as we shall see below, may well be the case of neo-corporatism.

## Democratization

Democratization in its most generic sense has involved the progressive extension of the citizenship principle to encompass a wider range of eligible participants and a wider scope of domains in which collective choice among equals (or their representatives) can make binding decisions upon all. Tocqueville was probably correct in identifying this as one of the strongest underlying trends in modern times. There have been, of course, some horrendous exceptions to this generalization, even within the heartland of democratic practice. It is, however, significant that even contemporary dictatorships usually define themselves as "régimes d'exception" and justify their existence on the grounds (however dubious) that they are dedicated to returning the polity eventually to some higher or more purified form of democracy. Moreover, recent developments in southern Europe and Latin America give us reason for renewed faith in the existence of such a *telos*, if only because tyrannical regimes seem incapable of resolving self-generated contradictions.[4]

Democracy has survived because its defenders have managed to assert its principle/procedures/institutions/qualities against the competing claims of other formulae for making collective choices and ensuring public order. Democratization has progressed because its promoters have succeeded in extending those properties into economic and social domains previously governed by other allocation rules and authority principles. But even where it has been successfully defended or extended, democracy has

always been compelled to make more-or-less enduring institutional compromises with a variety of antithetic practices. Social hierarchies based on tradition, market allocations of value, and specialized commands over knowledge are all intrinsically anti-democratic in the triple sense that they assign actors to unequal positions, they define certain domains as "inappropriate" for public scrutiny and accountability, and they resign citizens to the acceptance of choices made by more prestigeous, more knowledgeable, more powerful or better endowed others: notables, owners or experts. Nowhere among existing democracies has the principle of self-government among equals prevailed exclusively, and this is conveniently (if often misleadingly) symbolized by the insertion of a qualifying adjective before the term: e.g., representative, liberal, bourgeois, social, proletarian, people's, etc.

One reason for such standing compromises stems simply from the fact that contemporary democracies operate through a multiplicity of specific, historically-conditioned, political units – namely, secular, territorial states or, worse, large nation-states. Their pre-democratic heritage, their inveterate conflictualness, their sheer size and extension, their diffuse loyalties – all protect and encourage competing principles and make the democratic one more difficult to realize. Inherited status groups are respected for their past contributions; specialized elites are allowed to act with impunity on grounds of national defence or grandeur; representative agents are weakly connected to large and heterogeneous constituencies; nationalistic sentiments are manipulated in order to subordinate cultural diversities. If this were not enough, democracy itself has greatly increased the individual and group needs subject to collective choice and implementation and this, in turn, has resulted in an enormous expansion and centralization of the permanent apparatus for processing these claims, namely the State bureaucracy. The whole production of public goods becomes increasingly indirect and technically incomprehensible to the average citizen.

Marxists rightly have stressed for some time the limits placed on the democratic principle and its varying procedures/institutions/qualities by the existence of private property rights which protect their owners, constitutionally or otherwise, from popular scrutiny or control. The subsequent development of large business corporations with unprecedented resources and capabilities for affecting whole populations, and the internationalization of exchange and production relations which can effectively circumvent the control of even the most democratic of domestic polities have certainly added further elements of credibility to this critique. Nevertheless, the recent revival of interest in "State theory" has contributed surprisingly little to democratic theory – this, despite Marx's early and explicit recognition of a contradiction between the equality of citizenship and the inequality of ownership. From an instrumental or a functionalist

perspective, the Marxists' answer to the question of neo-corporatism's compatibility with democracy would seem to be so obvious as to be axiomatic. Given what is "known" about the distribution of "real" power in capitalist societies and bourgeois democracies, such institutional innovations can only be the product either of conscious intent by the ruling class to maintain its dominance, or of system constraints which ultimately must benefit the minority of owners over the majority of citizens. Indeed, one prominent approach to neo-corporatism explicitly defines the phenomenon in terms of its function of subordinating the working-class to the imperatives of capital.[5]

Recent discussions by Marxists in the conditions for establishing "class compromise"[6] and the role of "democratic antagonisms"[7] seem to be opening up new grounds within this tradition for a less aprioristic and, perhaps, less unfavourable treatment of neo-corporatism. Social democrats, of course, have been guilty, not only of harbouring, but of promoting such thoughts for some time,[8] although they have usually been prudent enough to keep the specific concept of corporatism disguised by euphemisms. What remains is for them to come more explicitly to terms with the limitations on democratic practice sketched above. These are imposed by the State form of political organization (especially in its increasingly centralized configuration) and by the technocratic basis of administrative power (especially in the realm of social welfare). With these elements in hand, we might then have an adequate theoretical basis for a critical evaluation of impact of neo-corporatism.

From this brief *Problemestellung*, we can draw the following conclusions:

(a) The citizenship principle provides the constant and common element to all theoretical efforts to define democracy but cannot be used alone to evaluate democratic performance. Nevertheless, if neo-corporatism can be clearly shown to diminish equal treatment or to reduce citizen acceptance of choices made by equal collective deliberation, then it must be regarded *prima facie* as undemocratic.

(b) Because this principle has been and can be expressed through a variety of decision rules and political institutions, to accept or reject neo-corporatist arrangements on the grounds of their compatibility with any particular configuration of them – even such established ones as majority rule or parliamentary sovereignty – would be ahistorical. That would be to deny one of democracy's greatest strengths, namely, its capacity for dynamic adaptation to changing circumstances. Nevertheless, neo-corporatism should be, at the very least, compatible with the prevailing "procedural minimum" – civic freedom, universal suffrage, etc. – since these are prerequisites for continued democratic experimentation with specific rules and institutions.

(c) The qualitative democratic relationships of participation, accessibi-

lity, responsiveness, accountability and competitiveness seem to provide us with the best criteria for evaluating neo-corporatist practices because they avoid the particularities and rigidities inherent in established rules and institutions and because they offer a more comprehensive perspective – individuals *and* authorities, input *and* output functions – from which to judge the whole or net impact of such practices.

(d) Existing democracies involve compromises with other ways of allocating values and exercising authority and are all, therefore, "second-best" approximations. A realistic and reconstructive critique of neo-corporatism should not be based on its presumed ability to eliminate completely these constraints – the State form of political organization, the competitive form of international relations, the administrative form of policy implementation, the commodity form of economic production – but on its observed capacity "to strike a better deal" when making compromises with them. From a dynamic point of view, the crucial question is whether neo-corporatism can better defend in the present and assert in the future the democratic principle of citizenship against competing ones than can other feasible configurations of decision rules or representative institutions.

## Neo-corporatism

The concept of corporatism, usually accompanied by some prefix such as "societal-", "liberal-", "bargained-" or, more recently, "neo-", burst upon the social science scene in 1974[9] and has since grown in prominence, to the point that it has been described by one author as "a growth industry".[10] Confused in political discourse with Fascism and authoritarian rule – not to mention with the French-Italian polemic usage which identifies it with the pursuit of narrow and immediate sectoral interests – and confounded in academic discussion by competing definitions and theoretical approaches, it has become a controversial subject, an "essentially contested" concept. Neo-corporatism (the *neo* is intended both to separate it from its historical predecessors – whether medieval or inter-war and to indicate its relative novelty) has been found everywhere – and nowhere. It has been credited with producing all sorts of goods – and charged with promoting all matter of evils. It has been described as an inexorable political trend – and called a passing academic fancy. Most of all, it has been difficult to define neo-corporatism clearly and consensually. One of its strengths has been its ability to speak to the concerns of scholars from different disciplines and orientations – each of whom, however, has tended to give this or her own twist to the concept. It has become virtually impossible to tell whether all of the contributors to this growth industry are talking about related, much less identical, phenomena.

For purposes of this essay, neo-corporatism refers to a recently emergent *political* arrangement – not to a new way of running the economy[11] or ordering the entire society.[12] It is concerned primarily with the activities of permanently organized and specialized *associations* – not units of production (firms, enterprises, corporations, etc.), not units of consumption (individuals, families, cooperatives, etc.), not units of status or affect (corps, colleagues, cliques, etc.) and not units of public authority (State agencies, ministries, parliaments, local governments, etc.). These associations seek to advance or defend interests by influencing and contesting collective choices. And they do this by intermediating between members and various interlocutors (mostly the State) without presenting candidates for electoral approval or accepting direct responsability for the formation of governments (i.e. they are not parties, caucuses, coalitions, etc). Any or all of the other above-mentioned units of political action may have a significant effect on the emergence or viability of neo-corporatism by supporting, opposing, or circumventing it, but they are not an integral part of its defining properties. Indeed, it is arguable that neo-corporatist practices have proven compatible with a rather wide range of surrounding units.

However, neo-corporatist arrangements are not the only way in which intermediation between interest associations and authoritative interlocutors can be institutionalized. Indeed, if one leaves aside medieval precedents and the short period of State-enforced corporatization under inter-war dictatorships, they are relatively recent and rare. For a considerable period, the predominant way of conceptualizing interest intermediation was "pluralism" and a very substantial and impressive literature on that topic was devoted to demonstrating that its arrangements were not just compatible with, but actively promotive of democracy. Much of the "cloud of suspicion" which hangs over neo-corporatism is due not just to the objections of those "utopians" who reject organized intermediation and incorporation of partial interests into policy-making on the grounds that it is destructive of the direct citizen role in public affair and of government by popular assembly, but to the suspicions of "realists" that these new arrangements represent a serious distortion and perversion of "proven" pluralist processes.

So, it would seem particularly appropriate in this context to define (neo-) corporatism by contrasting it with (neo-) pluralism. Such an approach has been criticized by some as epiphenomenal, as too "institutional" and "political-sciency", but this is to underestimate both the depth and breadth of the latter. For pluralism has never been "just" a theory of representation and political institutions. It also addresses important issues of the constitution of civil society, the nature of its component values and interest and the self-organization of groups within it, as well as of the formation of public policy and its impact on citizen satisfaction and

legitimation. It is no more "normative" in its bias than most, so-called, political economy models and since, in this case, its bias is unabashedly toward a particular type of "representative democracy", it provides us with a ready-made standard of evaluation.

Any mode of organized interest intermediation – pluralist, corporatist or, for that matter, syndicalist, monist, or whatever – must involve two contrasting aspects of the political process: (1) an arrangement for communicating and transforming member preferences into claims on others; (2) an arrangement for monitoring and influencing subsequent behaviour involved in realizing those claims. In the terminology of systems theory, it involves both input and output functions. These two aspects can be labelled: *representation* and *control*. Associations which operate in this political space are in the Janus-like position of intermediating between and focusing simultaneously upon two distinctive publics or clienteles. On the one hand, they must structure themselves internally and engage in such relevant activities that they offer sufficient incentives to their *members* so that they can extract adequate resources in the form of dues, fees, taxes, donations, voluntary labour, compliance, etc. to ensure at least their survival, if not their organizational growth. On the other hand, they must offer sufficient incentives to their *interlocutors* (i.e. State agencies, other associations, political parties) to be able to obtain other resources: recognition, toleration, access, protection, concessions, subsidization, etc., that are also needed for them survive and certainly to prosper. In other words, as specialized intermediaries, interest associations face a complex set of choices stemming from the often contradictory logics of appealing to members and exercising influence over interlocutors. Although pluralist theorizing about this domain has tended to underplay the control aspect and to make rather benevolent assumptions about the motives and activities of interlocutors (to the point of almost eliminating the distinctive coercive powers and administrative autonomy of the State), it is not difficult to reconstruct the more-or-less implicit assumptions about these elements of the political process and to contrast pluralist and corporatist conceptions of them.

Figure II outlines summarily the property space surrounding the role and resources of interest associations and contrasts the elements of a pure corporatist and a pure pluralist mode of structuring such relationships. The distinctions drawn are, needless to say, logical not empirical. They are subject to gradations and approximations, and no single existing polity exactly replicates the entire *Gestalt* formed by its dichotomous properties. More importantly, as a good deal of research has demonstrated, within any single polity the configuration of organized interests may vary considerably by policy arena. Even the same interest organization – business association, trade union or professional association – may be operating simultaneously in a more pluralist or corporatist fashion as it interacts with

**Figure II**

*Properties distinguishing pure pluralist and pure corporatist modes of intermediation*

| Associational role | I. *Representation (input)* | |
|---|---|---|
| Associational resources | Pluralist | Corporatist |
| I. *In relation to membres* | Multiple units | Monopolistic units |
| | Overlapping claims | Differentiated domains |
| | Autonomous interaction | Hierarchical coordination |
| | Voluntary adherence | Involuntary contribution |
| II. *In relation to interlocutors* | Mutual tolerance | Explicit recognition |
| | Opportunistic access | Structured incorporation |
| | Consultative access | Negotiative role |
| | Shifting alliances (log-rolling) | Stable compromises (package-dealing) |
| | II. *Control (output)* | |
| I. *In relation to members* | Persuasive conviction | Interest indoctrination |
| | Institutional (or leader) prestige | Organizational authority |
| | Discriminate treatment | Coercive sanctions |
| | Selective goods | Monopolistic goods |
| II. *In relation to interlocutors* | Provision of information | Organization of compliance |
| | Irresponsibility for decisions | Co-reponsibility for decisions |
| | Autonomous monitoring | Devolved implementation |
| | Mobilization of pressure (protest or disruption) | Withdrawal from concentration (secession) |

different agencies or levels of government. Much of the confusion in the literature stems from the propensity of scholars either to focus exclusively on representation rather than control or on relations with members rather than with interlocutors, or to "privilege" a particular issue arena and level of governance – one of the favourites has been national incomes policies

– and to generalize on the basis of the presence/absence, success/failure of such specific arrangements.

Figure II should make it clear – if nothing else – that neo-corporatism is a very complex bundle of changes in the relationship of associations to member interests and public policy-making. What is more, at a given moment in time and for a given issue in contention, the elements within the matrix may be far from consistent (as, incidentally, may also be the case with more pluralist configurations). For example, in a particular country or policy sector, associations may have a monopoly on representation, enter into hierarchical coordinative arrangements with each other and generate most of their resources involuntarily, but *not* have much control over the formation of member preferences, the formal authority to affect member behaviour or the effective sanctions to apply if their rules are transgressed. Even where these properties exist with respect to members, interlocutors (especially State authorities) may refuse to grant them corresponding corporatist "rights". They may *not* officially recognize existing monopolies, establish formal systems of guaranteed representation or depend on associational approval for the taking of policy measures. And even where they do this, when it comes to the actual allocation of goods or administration of regulations, they may *not* use associations to govern the compliance of affected interest, make associations co-responsible for subsequent decisions or devolve upon them authority to carry out directly the necessary tasks.

Seen this way, neo-corporatism is a very complex "bundle" of changes in the relation of organized interests to policy-making. Moreover, precisely because it has evolved in such a piecemeal and uneven manner, over a considerable period of time in response to very different circumstances in class relations, sectoral conflicts, status rivalries, international pressures, etc., and also perhaps because it has not been accompanied by a powerful and explicit ideology justifying its presence and homogenizing its practice, manifestations of corporatism are exceptionally diverse, difficult to capture, and subject to conflicting interpretation.

Let us set aside the considerable empirical problem posed by using such a logically consistent and abstract, ideal-typic, model to describe and analyse existing configurations. Contemporary research in Western Europe has been quite inventive in discovering and labelling mixed configurations with differing sets of participants at varying levels of policy-making.[13] What is especially relevant here is the speculative question of emergent properties. Are the polities of advanced industrial/capitalist societies moving inexorably in the same general direction away from pluralism toward corporatism – inhibited only momentarily and circumstantially by differences in historical points of departure, prevailing class conjunctures, international vulnerability, State capacities, political culture, etc? Or are they likely to remain frozen into different configurations

varying not only according to entrenched subjective national experiences but also with persistent objective differences in the structure of interest conflicts which define their particular classes, sectors, professions, strata, entitlement categories, or "causes"? For a tentative answer to this question, we must look now diachronically, not synchronically, into the historical processes which have promoted these changes in associability and policy-making – where they have occurred at all.

## Corporatization

Given the unevenness with which neo-corporatist arrangements are spread across countries and policy sectors and the variety in their structures, it may seem surprising that there are *any* general theories about how they may have come about, but there are. Moreover, there exist two strongly contrasting "hunches" which lead to virtually diametrically opposite evaluations with respect to democracy.

The one, particularly prominent in Scandinavia, tends to interpret neo-corporatism as part of a long-term and gradual historical process involving the rationalization of social relations. The emergence of monopolistic representation through the elimination of prior overlaps in functional domains or ideological programmes, the subordination of previously autonomous organizations to the hierarchical control of sectoral or class "peak associations", the granting and formalizing of direct access to State agencies and the acquisition of negotiative status, the acceptance of co-responsibility for courses of action subsequently decided upon, and the use of associations as agents for policy-implementation – all these hallmarks of neo-corporatism are regarded as an extension to the realm of interest intermediation of organizational developments which first became prevalent elsewhere. Moreover, since much of this simplification, specialization, "explicitization", standardization, bureaucratization, etc. of relationships is the result of policies carried out at the express will of the people, or, better, through its majoritarian partisan representative: social democracy, there can be little or no doubt about its impact upon democracy. Indeed, from this interpretive perspective, the historical development of bureaucracy and that of democracy go hand-in-hand. Their allocative and procedural principles may be different, even antagonistic, but they are intrinsically complementary. "Advances" in one of these choice/control mechanisms will call forth countervailing changes in the other. Neo-corporatism (or "societal bargaining" as Walter Korpi prefers to term it) is nothing but the inevitable, if not always intentional, side-product of such a dynamic process. As such, it is not merely compatible with democracy, but part of an advanced, better organized form of democracy in which capacities for collective action are more equally distributed

and relations to authority rendered more predicable and public.

The obverse "hunch" – also presented, of course, as a confirmed theoretical generalization – is that neo-corporatism is a product of "the laws of capitalist development", or more accurately, of a crisis in the historical course of that development. What is the nature of the specific crisis that induces such a change in organized class relations is a matter of some controversy among neo-Marxists, e.g. a "legitimation deficit", a productivity-profit squeeze, a change in the requisites of competitive survival in the world system, a shift in class forces, a decline in bourgeois parliamentarism. But what is clear from this perspective is that neo-corporatism is a deliberate strategy on behalf of a propertied minority enjoying unequal benefits – not an unintended side-product of a majoritarian demand for more equal treatment. The benefited class, i.e. capitalists, may not struggle explicitly to establish such institutions; Indeed, they may stubbornly cling to their out-moded ideology of individualistic competition and voluntaristic cooperation, but "their state" functions predicably and reliably to ensure their social reproduction and, therefore, intervenes to protect their long-term, objective interests by promoting the formation of neo-corporatist arrangements. Since these will not operate without some degree of collaboration from "social partners", trade union leaders, consumer representatives, and so forth must be hegemonically convinced by internalizing the norms of reformist ideology and/or personally coopted by enjoying organizational payoffs to participate in them. Neo-corporatism is, therefore, intrinsically undemocratic, a major barrier to the attainment of a more participatory, just and responsive polity. Fortunately, however, according to these theorists it can only be a temporary expedient. Its dependence on cyclically-generated surpluses, its asymmetric distribution of benefits and its ultimate inability to master the contradictions of capitalist production will destroy it in the longer run. Whether it will be replaced by an advancement to a higher form of democratic socialism or recession to a lower form of authoritarian capitalism depends on "the balance of class forces" at the moment this intrinsic crisis manifests itself – and it is the function of radical democratic theory to promote the more desirable outcome.

Between these sharply divergent poles of analysis and evaluation lies a somewhat confused and eclectic bunch of interpretations of why neo-corporatism has emerged, who is likely to benefit from its persistence and what will be its probable future, while acknowledging the broad, parametric effect of the "organizational revolution"[14] which swept contagiously across countries and social groups since roughly the 1890s, and the shift in internal class and external competitive relations which has occurred in advanced capitalist societies especially during the protracted, post-World War II expansion, these eclectic approaches tend to stress both the differential spread of the organizational phenomenon and its relevance to

internal conflicts not strictly determined by the struggle between capital and labour. Neo-corporatism becomes less the predictable or inevitable product of some general trend than an opportunistic, pragmatic and often unintended response to disparate circumstances. It has been promoted (and occasionally inhibited) by episodic events such as wartime exigencies, peacetime reconstruction, foreign occupation and regime discontinuities; it has been affected by constant national differences in size, strategic location, resource bases, patterns of ethnic, linguistic and religious diversity, etc.; and perhaps most importantly, its emergence has depended on specifically political variables such as the electoral strength of Social Democracy, the centralization of administrative structures, the distinctiveness of recruitment and training of civil servants, the rootedness of territorial-parliamentary systems of representation, the ideological fragmentation of class forces, and the shifting substantive content of public policy. These are conditions which tend to vary much more across and within countries than do their respective levels of organizational rationalization or capitalist development. However, by introducing so much contingency into the pattern of its emergence and functioning, such eclectic approaches to neo-corporatism make much more difficult the evaluation of its positive or negative consequences for democracy. Their answer (and mine) is likely to be a limp and inconclusive "it depends".

## Associability and the pursuit of interests

The *genus* of which neo-corporatism is such a relatively recent species is associability – the propensity for groups of persons within a larger polity to join together in some more-or-less formalized way to pursue through collective action interests they believe they have in common. Political authorities are not always tolerant of such efforts. Indeed, it would be more accurate to say that, historically, they have most often been outright hostile to self-constituted, self-regarding and self-actuating "factions" defining them as subversive of public order and divisive of State sovereignty. This "worms-in-the-entrails-of-the-body-politic" view has recently resurfaced in the conservative literature on "overload" which claims that the alleged ungovernability of contemporary democracies is due to excessive activity by groups acting in selfish disregard for higher public purpose and usurping tasks which should be left to enlightened elites.[15]

Nor have democratic theorists and political progressives always been supportive of associability. The sheer presence of such intermediaries could be seen as a factor distorting the relation between individual citizens and sovereign authorities – interjecting partial wills into the process of general will formation or protecting reactionary factions against the tide of historical transformation. This theme of direct, unmediated democracy was articulated most emphatically by Rousseau and reached its apotheosis

during the French Revolution in the Loi le Chapelier of 1791 banning "all types of associations among citizens of the same status or profession under any pretext or any form that exists". It returned again, over a century later, this time as farce, in the pretence within "popular democratic" regimes that the ruling revolutionary party represents all legitimate interests and that whatever associations are tolerated must be subordinate to its will.

Elsewhere the evolving practice of democracy began to recognize freedom of association as a basic civic liberty – part of the procedural minimum we discussed above. Beginning with the constitutional breakthroughs of the United States of America (1792), Belgium (1830), Switzerland and Denmark (1848-9), it became increasingly difficult to deny citizens the right to found or join associations on account of their class, race, condition of dependence, sex or age. Authorities did (and still do) use other legal pretexts to prevent certain groups from assembling and, most of all from acting collectively, e.g. "subversive intent", "restraint of trade", "breach of the peace", violation of property". They also can employ other, less legal, impediments such as police harassment, denial of access to public facilities, tax penalties and outright violence, but within most social groups in Western Europe and North America citizens have been able for some time to create permanent organizations for the defence of their interests; recruit (and even reject) members as they see fit; establish their own institutions of self-governance; acquire and accumulate property; sue (and be sued by others); engage in the activities of their own choosing; petition authorities when they please on the topics they prefer – subject only to certain modest requirements of formal registration and respect for the criminal law and tax code. More recently and less widely, associations have acquired the legal right to exclusive recognition as *bona fide* negotiators for their members, to be consulted before decisions within their interest domain can be made, to receive information and subsidies from the State officials, and to participate directly in the implementation of public policy.

Most of these rights and attendant obligations were acquired by associations through struggle, compromise and pragmatic sanction, punctuated only occasionally by constitutional revision or landmark court decisions. Democratic theory only belatedly (and in some cases, reluctantly – cf. Madison's famous 10th Federalist paper) caught up with these changes, although Alexis de Tocqueville had the genius to recognize the significance of associability at a relatively early stage of its development in what was then its most advanced site: the United States. His famous hypothesis has stood the test of time astonishingly well: "For men to become or remain civilized, the art of association must develop and perfect itself among them in the same measure as the equality of conditions among them grows."[16] From this perspective, one can deduce that the more freedom of action

associations can enjoy, the more members they can recruit, the more activities they can engage in, the greater the coverage they can provide for citizen interests, the more democratic the polity is likely to be, to become or to remain. Leaving aside that *démocratie* is not exactly what Tocqueville meant by *civilisation*, this postulated linkage between associability and democracy seems to be one of the closest, most virtuous and best established we have – rivalled only by that between elections and democracy. How could one possibly attack such an assumption?

Robert Michels probably fired the first salvo when he pointed to the resources which incumbent leaders within associations could manipulate in order to maintain oligarchic control.[17] The subsequent increase in size with the emergence of large-scale national organizations has entailed an inevitable loss of personal contact among members and a more indirect system of internal governance. As one distinguished French student of associability put it: "associations are subject to the same processes of growth and giantism as private enterprises."[18] All those imputed benevolent qualities of Tocqueville's which hinged on social intimacy and democratic experience became more dubious. Empirical inquiry subsequently discovered that only exceptional associations – even in relatively pluralist contexts – were able to sustain the sort of active participation, open competition and close accountability which made them into "schools of civic virtue".[19] The demonstrated prevalence of entrenched oligarchy also raised the uncomfortable possibility that association leaders might be collecting "political rents" for themselves in the form either of self-serving appropriation of group benefits or the public promotion of interests not shared by members.

The next serious general assaults on the associability-democracy citadel came from Mancur Olson in his *Logic of collective action.*[20] Working within a voluntaristic and individualistic frame of reference, he argued that limitations on the ability to extract contributions for the provision of public goods would encourage associational entrepreneurs to offer increasing quantities of selective goods whose consumption could be restricted to members. In effect, interest groups would have to become more like business firms if they wished to recruit members and develop their resources further. This would distort their internal processes away from the democratic concern with mobilization and representation, thereby, diminishing the role of militant action, ideological conviction and voluntary participation in favour of passive conformity, material consumption and professional management. It would also tend to free association officials to pursue public goals not necessarily shared by members, thereby, increasing further the rate at which "political rents" were likely to be levied.

Olson's perspective also challenged the validity of one of the most central (and democratic) assertions of pluralist theory – namely, that

given an equal freedom to associate and an equivalent intensity of interest, a spontaneous and dynamic process of countervalence would occur and, thereby, prevent any entrenched elite or stable coalition from dominating the policy process. Because the propensity for voluntarily contributing to collective action is unevenly distributed across interest categories of differing sizes and degrees of dispersion, "privileged groups" would naturally emerge and large categories of interests would remain effectively latent – protected only by their episodic voting power or the benevolent intervention of elites on their behalf. Olson barely refers to the class bias intrinsic in such disparities, but Claus Offe and Helmut Wiesenthal have developed this theme much more coherently and extensively in a recent article.[21]

But these critiques of the free associability-political democracy nexus are, in a sense, "pre-corporatist". Convinced pluralists may recognize them as distortions or problems, but they treat them as "améliorable" by legal reform or "avoidable" through political struggle. After all, under-organized latent interests do have access to alternative channels of expression, e.g., territorial and partisan representation or single issue movements. Other, "solidarity" and "purposive", incentives can be used to recruit members and attract contributions.[22] Individuals can still "vote with their feet" against unrepresentative leaders, and shift to alternative suppliers of desired goods and services.[23] Competition between overlapping associations can serve as a functional substitute for internal oligarchy. Mobilization of latent groups by aspiring entrepreneurs or conspiring elites can serve as a check on dominant "power blocs".

## Neo-corporatism and the science of organization

Above we have noted with approval Tocqueville's argument that promoting "the art of association" was a necessity if democracy is to keep pace with social transformation. And neo-corporatist arrangements do seem to have a significant and independent effect upon the properties of participating interest associations. Their densities of membership go up almost to the point of saturation; their organizational resources increase and become more equally distributed across incorporated interests; their range of activities becomes more varied and authoritative; their scope of representation becomes more comprehensive. Since these arrangements have not been imposed by an authoritarian regime and asymmetrically manipulated from above to favour some interests over others – as was the case of inter-war State corporatist experiences – and have been freely chosen, voluntarily maintained and occasionally renounced by organized groups seeking access to public authority, they would seem to embody the sort of adjustment that Tocqueville advocated. Moreover, again in contrast to the inter-war period, the emergence of such arrangements was neither envis-

aged as a replacement for traditional liberal forms of territorial representation, nor have they proven incompatible with such hallmarks of democratic procedure as freedom of speech and press, respect for the constitution, electoral competition, public disclosure, etc. In factor, many of the countries scoring highest in various indicators of neo-corporatism: Sweden, Norway, Austria, the Netherlands, for example, have been in the vanguard of extending equal citizen rights into new domains and equalizing the benefits from public policies.

What Tocqueville could not anticipate and what his subsequent pluralist epigones failed to recognize is that the development of permanent, specialized and professionalized intermediaries between citizen and State might transform "the art of association" into "a science of organization". Instead of merely *re*-presenting independently formed member preferences, associations could become institutions for inculcating and managing member interests. Instead of remaining multiple "vehicules for meaningful participation", they might increasingly become monopolistic "providers of indispensible services". What is more, instead of providing "havens of protection" for privately enjoyed satisfactions, they can come to constitute more and more "conduits of intervention" for publicly supplied coordinations.[24] Tocqueville assumed that voluntary associations would provide lively, significant and alternative sources of individual identification and political experience – a sort of "secondary citizenship" outside the official one – and that their activities would lessen rather than increase the need for involuntary, authoritative coordination of individual behaviours within society. Neo-corporatism has the potentiality of altering the relationship of members to their associations and the role of the latter inside the official realm of authority. In other words, it might be undermining the "virtuous circle" between associability and democracy.

Of course, all forms of specialized and insitutionalized interest polities violate at least two prevailing norms of democracy. As several analysts have pointed out,[25] *votes are counted* – presumably equally if the elections are honestly conducted – but *interests are weighed* – presumably proportionately in relation to their functional importance or political clout. Moreover, in voting systems decisions are usually taken by forging a consensual compromise to which all actors agree – however reluctantly. The historical emergence of this mode of structuring political exchanges can be interpreted as an effort by *minorities* who wished their interests weighed and who were aware that their compliance was necessary for collective choice to be effective to protect themselves against *majorities* who wished their votes counted and who were confident that they could overcome dissenters by force of numbers. Interest politicians successfully pitted one procedural norm (freedom of association) against another (freedom of electoral expression) – and one decision rule (consensus) against another (majoritarianism).

All neo-corporatism has done is to take this differentiation within democratic polities much further. With its hierarchic coordination, State recognition, policy concertation and devolved authority, it has separated more effectively than before the interaction between interest associations and public authorities from the vagaries of electoral success and legislative choice. With its monopolistic representation, elimination of overlapping domains and quasi-voluntary membership, it has restructured the conditions of competition among associations by raising the cost of exercising "voice" through alternative channels and foreclosing the possibility of resorting to "exit". Without easy access to other channels of interest expression, effective choice between competing organizations or a reasonable chance to create rival ones, individual citizens in neo-corporatist systems may find themselves deprived of the resources necessary to hold "their" representatives accountable and to ensure that policies pursued will be responsive to "their" concerns.

Presumably, this places a much more substantial burden on the quality of the political process *within* neo-corporatist associations than was the case in more pluralist systems with their overlapping structures and ease of entry and exit. The question becomes whether competition and participation are likely to be greater within monopolistic, heteronomous organizations that within multiple, autonomous ones. There is every reason to expect the contrary since – with compulsory membership, provision of indispensible services, public subsidization, devolved authority and/or mandatory licensing – neo-corporatist leaders can be largely freed from the constraint of inducing members to contribute and, hence, from the need to conform closely to members' preferences in order to extract those contributions. Moreover, their symbiotic relation to State authority encourages, if not requires, them to adopt a longer-term view based on technical calculation and expert opinion. This enhances the role of professional staff at the expense of voluntary labour and more "spontaneous" expressions of preference. Therefore, one can presume that, in the normal course of transacting interest business, leaders will tend to discourage member participation and to underplay the role of militant mobilization. Followers will learn that associations have less and less to offer in the way of affective interaction and collective solidarity. Whatever aspirations members still have for "secondary citizenship", they are likely to take elsewhere, to political parties, community action groups, single issue movements, if they do repress them altogether.

Which is not to say that neo-corporatist practices in representation and policy-making are intrinsically antil or undemocratic. It does suggest that, where they have emerged and become entrenched, they may be altering the quality of democracy. Their "science of organization" has already changed its decision rules and procedural norms.

## Neo-corporatist practices and the citizenship principle

In our effort to develop normative standards for evaluating its performance, we defined modern, representative democracy as a general principle in search of a certain qualitative relationship between rulers and ruled. Connecting the two are decision rules, procedural norms and political institutions which have varied considerably over time – no matter how established and definitive they may appear at a given moment. Except in instances of dramatic re-founding after the collapse of an authoritarian regime or in periods of deliberate reform in the face of manifest crisis, these specific rules, procedures and institutions tend to change slowly, often imperceptibly and usually by consensus under the pressure of opportunistic situations, international diffusion, or evolutionary trends.

Neo-corporatism is a good example of such a transformation within democratic polities in which the "procedural minimum" has been respected, but substantial changes in such things as majority rule, parliamentary sovereignty, public deliberation, etc. have occurred. Perhaps precisely because these contemporary trends in the organizational structure of interests and in their relation to policy-making have not been backed by an explicit ideology – again in contrast to inter-war, State or authoritarian corporatism – and because they evolved in such a piecemeal, uneven and almost surreptitious manner within distinct policy arenas, they have largely escaped evaluative scrutiny. Only once they have accumulated over time, so-to-speak, and are manifestly affecting a wide range of producer and consumer, as well as citizen, behaviours is the question of their compatibility with democracy likely to arise – and to attract the attention of scholars as well as activists.

Our first and most diffuse evaluative standard is whether these arrangements violate the principle of citizenship. Do they diminish the extent to which individuals have equal opportunities to act as citizens and to be treated as equals by their fellow citizens? Do they reduce the extent to which citizens feel obligated to respect choices made by collective deliberation among equals (or their representatives)?

If one equates the opportunity to act as citizen only with voting and the obligation to conform only to laws which have been certified by a sovereign legislature, then neo-corporatism is manifestly contrary to the citizenship principle. It introduces elements of "weighted" calculation and consensual bargaining with privileged minorities which clearly violate the sacred norms of "one man, one vote" and "the most votes win". It generates binding commitments which either are never subject to parliamentary approval or involve a mere officializing of package-deals hammered out elsewhere. However, if one broadens the notion of equal political opportunity and treatment to include intra-electoral periods and

extra-electoral processes, then neo-corporatism can be interpreted as extending the citizenship principle.

Basically, what it does is to resolve "the paradox of liberal associability" i.e. the fact that where the freedom to associate is equally accorded but the capacity to exercise this freedom is unequally distributed, those that most need to act collectively in defence of their interests are the least likely to be able to do so. Small, compact and privileged groups who are already better able to advance their interests through existing economic and social exchanges than larger, more dispersed and equally endowed ones, will find it easier to recruit members and extract contributions for a further defence of their interests in the political realm – if and when such a response is required. Hence comes the theme of the "institutionalization of bias" in pressure politics which has been decried by so many critics of pluralism.[26] What neo-corporatism does is shift the basis of associability from a predominantly voluntaristic and individualistic calculus to one where contributions become more generally binding on all members of a relevant category (or more difficult to avoid), and mutual recognition and offical certification protect the role of specific collectivities at the expense of competing fragments or individual actors. In short, as "free-riding" and "free-booting" become increasingly difficult under such arrangements, virtually everyone can be made to contribute and conform to associative action. This can have the effect of evening out considerably the organizational capacity of competing groups, particularly capital and labour. In addition, most neo-corporatist forums are based on highly formalized systems of parity in representation and, not infrequently, produce policies which make the participating "social partners" co-responsible for their implementation. Under such conditions, organized socio-economic interests may never be equally counted, but they are likely to be more equally weighed than they would be if citizens invested voluntarily and individually their own disparate resources and personal intensities in the liberal "art of association".

The normative problem with applying this "science of organization" to interest intermediation is that it may make more equal the capacities for exerting influence of incorporated collectivities, at the same time it purposively excludes others which may be affected by their deliberations. So far, neo-corporatism has privileged interest organized along functional lines of production within a capitalist economy – classes, sectors and professions. Its relative success has depended on restricting the number and identity of participants and passing on the costs to those not directly represented in its deliberations: consumers, taxpayers, youths, feminists, irregular workers, foreigners, cultural minorities, nature-lovers, pedestrians, prohibitionists, etc. Granted that the more comprehensive scope of the associations engaged in neo-corporatist bargaining may encourage them to take into account some of these "marginal" interests, for example, when a

comprehensive trade union calculates the effect of its demands on its member interests as consumers or when a national business association agrees to moderate its position in deference to the need for environmental protection,[27] but this is a tenuous and contingent relationship – hardly reliable enough in the long-run to lead to an effective equalization of influence for such categories of citizens. Existing corporatist associations which defer too much to such interests risk a paralysis of their own internal decision structures and/or a defection of their own core supporters.

One "democratic" answer would be to extend the process of corporatization to cover *interests* structured along distributional lines or *causes* generated by cultural and ideological diversity, but that hardly seems feasible. Establishing monopolistic, hierarchically coordinated and topically differentiated national associations for, say, consumers, taxpayers, youths, environmentalists and foreign residents would likely involve such extensive State intervention and subsidization that it would be difficult to avoid the appearance, not to mention the reality, of manipulation and cooptation from above. The officially recognized associations would be simply disavowed by their nominal members and lose all credibility for contracting in their names. In addition to which, many of these groups define their very existence in ways that defy professionalized representation and bureaucratic *encadrement*. To be organized corporatistically would destroy the very basis of their collective identity. Finally, even if the organizational problem could be solved, bringing such a quantity and variety of recognized interlocutors into the policy-making process on a co-equal basis would destroy the properties of small group interaction, specialized competence, reciprocal trust and propensity for compromise which have contributed so much to the viability of existing neo-corporatist arrangements.

In summary, a pattern of more equalized and formally structured exchange among associations has emerged in some democratic countries – a sort of *corporatism for the functionally privileged* – which could be defended as a direct extension of the citizenship principle outside the electoral-parliamentary arena in ways that go beyond the formalistic opportunities afforded by liberal associability. Moreover, its operation has undoubtedly had the indirect effect of promoting policies which have extended citizen rights to protection against unemployment, to more extensive welfare services and to representation within institutions previously governed by other authority principles, especially business firms and State agencies. Citizens of pluralistically structured policies have suffered significantly greater inequalities in all these domains. Its unintended consequence, however, has been to consolidate a disparity between these more equally *compétent* and privileged class, sectoral and professional interests and less equally *compétent* and organized ones – leaving a sort of

*residual pluralism for the distributionally disadvantaged and the culturally underprivileged.* Since there appear to be serious impedimants to extending neo-corporatist practices to these latter interest domains and since at least some of these appear to be of genuine concern to the citizenry of contemporary democracies, neo-corporatism is neither fully compatible with the citizenship principle nor are decisions made under its auspices likely to go unchallenged by those who are expected to obey them. However, like so many of its forerunners in the history of democratic development, its norms and institutions may long be tolerated as a second best compromise: "a better system than those that preceded it and those that have hitherto followed it".[28]

## A shift in the quality of democracy?

By now, it should be fairly evident what neo-corporatism can do to the quality of democracy – at least from the evaluative perspective adopted in this essay. The reader will recall Figure I in which five qualities were generated from a matrix which distinguished dichotomously between two units of reference (individual citizens or public authorities) and two aspects of governance (openness to inputs "of the people" or processing of outputs "for the people"), and which placed one feature of democratic performance at the centre, presumably linking all the other virtues in a coherent whole.

Expressed in its graphic terms, neo-corporatist arrangements shift performance to the right – away from a concern with *participation* and *accessibility* toward a greater emphasis on *accountability*, and *responsiveness*. Individual citizens become less intensely and directly involved in political life; at the same time, organizations active in their interests become increasingly integral components of the policy process. The number and type of interlocutors with equivalent and effective access to authorities decrease considerably due to the recognition of monopolies, the creation of associational hierarchies and the formalization of functional systems of representation; at the same time those that are able to obtain such privileged status acquire more resources and become more indispensible to the management of public affairs so that arbitrary (and often self-serving) actions by State officials become less likely. Subjects of collective choice which were highly politicized, i.e. subject to intense citizen concern, public debate, group mobilization and extensive pressure become less so; at the same time that institutions of administrative and market allocation which were previously defined as outside the realm of democratic polities become subject to greater scrutiny by political associations.

In the midst of this shift from participation/accessibility to accountability/responsiveness lies a phenomenon which modern democratic theory

has been ill-prepared to analyse or even to recognize – namely, the development of "private" or "class governance". Perhaps, one major reason for this stems from its historical roots in liberalism. Democratic theory was originally closely identified with the liberal struggle against the constraints that obligatory associability had placed on the economic and social behaviour of individuals: guild restrictions, State-chartered monopolies, licensing provisions, etc. It continued to regard all associations which subsequently grew up under its tolerance and encouragement as purely voluntary and autonomous, an embodiment of that original resistance to regimentation and loss of individual freedom.

Neo-corporatism changes not merely the resources of associations and the nature of policy-making. It can also radically alter the relationship between interest groups and their members. Instead of merely aggregating independently formed preferences and articulating them before authorities, its associations acquire an enhanced capacity for defining the interests of members and controlling their behaviour. In the pluralist idiom, *information* is the key resource involved in interest exchange; in the corporatist mode, it becomes *compliance*. Associations do not just inform policy-makers about the intensities of preference and likely reactions of their members, expecting officials to react accordingly; they also agree – for a price – to deliver member compliance to contracts negotiated with the approval of public authorities. All this presumes, of course, that it will be to the long-term benefit of members to be forced to cooperate irrespective of their individualistic, short-term preferences. Occasionally and always reluctantly, neo-corporatist organizations may even have to wield directly the coercive powers necessary to keep dissident members (and even nonmembers where contracts are extended to cover a whole category) in line: fines, expulsions, refusal to provide services, loss of licence, etc. Whether this authority is generated consensually from within or devolved legally upon it from without, the net result is the same. The society acquires a set of parallel institutions of semi-private or semi-public governance capable of coordinating the behaviour of some large social aggregates – classes, sectors, professions – without directly burdening or involving State authorities. This may provide one element for explaining why the more neo-corporatist polities have proven demonstrably more "governable" in recent decades than pluralist ones at a similar level of capitalist development and organizational complexity.[29]

This leaves us with competitiveness, the quality of democracy which putatively ties all the others together. What happens to it under the auspices of neo-corporatism? Obviously, some forms are eliminated altogether or reduced to insignificance. Groups with overlapping domains no longer compete for members or for access to public authorities on the same issues. Factions within associations are less likely to risk investing their resources in founding alternative organizations, if only because other

public and private interlocutors will persist in recognizing only the officially monopolistic one. Under a general process of incorporation, highly specialized or very particularistic, "maverick groups" will find it increasingly prudent to merge with larger and more established units or to accept coordination from overarching "peak associations" if they do not wish to suffer a progressive marginalization from the policy-process.

It is not clear whether the politics *within* neo-corporatist organizations is likely to become more competitive as that between them diminishes.[30] Certainly, the rewards for winning office become more substantial with the increase in associational resources and semi-public functions, but that may only encourage entrenched oligarchies to defend their positions more assiduously and tempt State and party officials to intervene in order to ensure that interest interlocutors will continue to play the responsible and respectful role assigned to them. At the level of national peak associations, executive leaders become highly visible and influential figures who can count on help from "outsiders" provided they agree to stay within the rules of the corporatist game.

But this does not mean that competitiveness disappears altogether under such arrangements. Rather its effect tends to become more implicit than explicit, more potential than observable. One must never forget that neo-corporatism is a chosen, not an imposed, strategy for the promotion and defence of interests, and that it is not the only mode of intermediation between citizens and authorities. Associations can withdraw from negotiations patterned this way – and they can survive, even prosper, by engaging in classic pressure politics. Specific issues can be taken to other arenas – and they can be articulated through single issue movements or spontaneous protest actions. Association members are also voting citizens – and they can express their dissatisfaction by switching allegiance among existing parties or by supporting new ones. Parties which have promoted corporatist arrangements can lose elections – and their successors in government may choose to dismantle or ignore those arrangements. Parliaments can assert their legal sovereignty – and they can refuse to ratify the social contracts which are put before them. Members can refuse to obey the directives of their associations – and the sanctions available may be sufficiently weak or difficult to wield that they can get away with such defections. The fact that such occurences have been relatively rare in neo-corporatist polities does not alter the latent role that competitiveness continues to play in setting boundaries upon such arrangements. Participants in them are forced to anticipate that such reactions could occur and to adjust their bargaining behaviour accordingly. They cannot act as if neo-corporatism were the only game in town. Observers, however, who predicted its imminent demise after each wildcat strike or electoral failure of Social Democracy have been generally disappointed. It even appears to be surviving under the conditions of increased national and international

competitiveness induced by protracted recession and consequent failure to meet such performance goals as full employment and economic growth. Nevertheless, the politics of these countries has not settled into some "post-problematic" consensus. Controversial items still manage to get on their agendas for collective deliberation; citizens continue to be offered real choices; associational leaders know they must be accountable to member preferences – just as they know they must be responsive to system imperatives.

Moreover, one cannot ignore the fact that many of the polities where neo-corporatist practices have become most firmly entrenched have either inherited (e.g. Switzerland) or experimented with (e.g. Sweden, Norway, Denmark, Austria, Western Germany) a wide range of institutional innovations which have extended the equal rights of individual citizens in their direct interaction with public officials and partisan representatives: referenda, proportional representation, ombudsman systems, subsidies for political parties and citizen groups, elections to works councils, public disclosure laws, decentralized administration, protection of personal data, profit sharing arrangements and so forth. One can argue that not all of these have had that much of an impact (and some have been very selectively implemented), but one can hardly fault these systems for not trying. It is at least plausible that discussion of them and their eventual presence in political life has effectively compensated for some of the more insidious and less positive effects that creeping neo-corporatism has had upon the other democratic qualities of participation, accessibility and competitiveness.

Finally, one must acknowledge the almost complete absence of popular resistance to neo-corporatist trends in those countries. This is all the more remarkable since they have rarely been defended explicitly and globally. Ordinarily they have only been sanctioned pragmaticaly on a case-by-case basis. No one confesses to being a "Corporatist", or even to some euphemism thereof. There exists no explicit justification of its practice in terms of its conformity to democratic principles or procedures. And yet citizens have by-and-large accepted it upon reflection. They might recognize that its emergence has altered their rights and obligations and they might occasionally grumble to survey researchers that "organized business", "organized labour" or "organized professions" seem to have too much influence, but few if any seem to feel that they have lost more than they have gained by entrusting the management of their interest politics to such intermediaries. Ironically, it is in those countries whose interest associations have been least corporatized that one most often hears the epithet: "Corporatist" thrown at opponents, and that intellectuals denounce all signs of its prospective emergence as a threat to traditional freedoms and democratic institutions.

## The spectre of vicarious democracy?

So, does this mean that the more neo-corporatist polities are already headed toward some new form of post-individualistic, vicarious democracy, with other advanced industrial/capitalist societies soon to follow? That the famous myth of the rational, well-informed and active citizen has finlly been put to rest and been replaced by the spectre of the reasonable, well-staffed and recognized association as the basic unit of democracy? That the notion of a civil society composed of natural groups volutarily entering into exchanges in the pursuit of their own autonomously defined preferences and capable of reproducing itself without the constant intromission of the State has given way to a vision of a semi-public society composed of artifactual organizations compulsorily negotiating compromises in the pursuit of their members' imputed interest and capable of sustaining itself only by symbiotic interdependence with public authorities?

Let us leave aside the probability that neo-corporatism and vicarious democracy may well be a solution to the problem of modern interest conflict confined to particular countries and national circumstances. Small size, high international vulnerability, well-established State legitimacy, centralized administrative structures, clear preponderence of class cleavage over other bases of social and cultural conflict, ideological hegemony of social democratic over liberal bourgeois values are all factors which seem to have contributed to the emergence of such a pattern, although they may not necessarily all be prerequisites for such an outcome in the future.

I suspect that the answer to the "paradox of corporatist associability" – to its ambiguous impact on the *practice* of democracy – eventually lies in the truth of what is one of the most central tenets of the *theory* of democracy, namely, that for a polity to be really responsive to the needs and concerns of its citizens, these individual citizens must *participate* actively and freely in the definition of those needs and the expression of those concerns. They must not only have the "enlightened understanding" of their interests which Robert Dahl so rightly stressed, but they must also have the resources *and* the desire to engage in the political struggle necessary to make sure their preferences are taken into consideration either by those who govern or by seeking themselves to govern. Specialized experts, organic intellectuals, designated spokespersons, professional intermediaries, benevolent rulers, etc. may, in some contexts and for some period of time, be better informed and more capable of interpreting the interests of social groups, but unless they are kept accountable by an active citizenry, their theories and suppositions about what is good for their members, clients, followers, etc. are likely to prove erroneous in the long

run. What is more, the organizational and political "rents" which these intermediaries extract for the service they perform will systematically distort the very content of demands made upon the polity.

The progressive assertion of interest politics, its conversion from an "art of association", into a "science of organization", may have greatly changed the identity of relevant actors. It may have expanded the resources and extended the range of such intermediaries. The emergence of a neo-corporatist mode may have increased the immediate governability, improved the aggregate economic performance and equalized access to policy-making in advanced capitalist societies, but the "vicarious democracy" which has accompanied these transformations may not prove so satisfying and in the long run so governable. *Rules may become more accountable under such arrangements, but to the wrong collectivities* – not necessarily to the units with which persons voluntarily identify and from which they naturally derive a sense of shared existence, but to those which struggle, convenience, connivance and luck have allowed to become formally organized, often at levels of aggregation far above that which would have been spontaneously forthcoming. *Governments may also be more responsive, but to the wrong needs* – not necessarily to those which individuals would themselves feel and become concerned about, but to those which professional intermediaries have defined and promoted as the "real" interests of their respective memberships or clienteles – often while including substantial sidepayments for themselves and the organizations which they control.

Whatever impact the organization of interest politics has had upon political performance, whatever has been the relationship between neo-corporatism and governability, whatever both have done to growth, equality and democracy, it is difficult to imagine that these changes have completely voided the old liberal adage that "each individual person is the best judge of his or her own interests". Ultimately, if not immediately, the polity will be judged by its ability to satisfy *these* interests – not just *those* which have been identified, given generic labels and packaged collectively by intermediaries and to which authorities have presumably been dutifully accountable and responsive. Moreover if, among these "really felt" interests of individual citizens, are distinctively political needs for active participation and close access to rulers, then one would have even more grounds for suspecting that the sort of "vicarious democracy" promoted by neo-corporatism will prove to be but a passing phase – hopefully, an appropriate and proportionate (if temporary) adjustment in the "art of association" that Tocqueville thought was so necessary to keeping our collective existence "civilized" while our individual conditions were becoming "equalized".

## Notes

[1] Stanley Hoffmann, "Some notes on democratic theory and practice", *The Tocqueville Review*, Vol. II, No 1 (winter 1980), pp. 59-76.

[2] Norberto Bobbio, "Are there alternatives to representative democracy?", *Telos*, No 35 (spring 1978).

[3] Robert Dahl, "On removing certain impediments to democracy in the United States", *Political Science Quarterly*, Vol. 92, No 1 (spring 1977), pp. 10-13.

[4] Guillermo O'Donnell & Philippe C. Schmitter, "Political life after authoritarian rule: tentative conclusions about tentative transitions", in *ibidem* (eds.) *The Transition from Authoritarian Rule: The Prospects for Democracy in Latin America and Southern Europe*, (Baltimore: Johns Hopkins University Press, forthcoming).

[5] Leo Panitch, "The development of corporatism in liberal democracies", in P. Schmitter & G. Lehmbruch (eds.), *Trends Toward Corporatist Intermediation*, (London: Sage Publications, 1979), pp. 119-146.

[6] Adam Przeworski, "Material bases of consent: politics and economics in a hegemonic system", *Political Power and Social Theory*, Vol. 1 (1980), pp. 23-68.

[7] Ernesto Laclau, "Democratic antagonisms and the capitalist State", in David Robertson & Michael Freeman (eds.), *Frontiers of Political Theory*, (The Harvester Press, Brighton, 1980), pp. 101-139.

[8] Cf. Leo Panitch, *Social Democracy and Industrial Militancy*, (Cambridge, England: Cambridge University Press, 1976); Francis G. Castles, *The Social Democratic Image of Society*, (London: Routledge & Kegan Paul, 1978); John D. Stephens, *The Transition from Capitalism to Socialism* (Atlantic Highlands, No 7: Homanities Press, 1980); Walter Korpi, *The Democratic Class Struggle*, (London: Rontledge & Kegan Paul, 1983).

[9] Gerhard Lehmbruch, "Consociational democracy, class conflict and the new corporatism", in P. Schmitter & G. Lehmbruch (eds.), *op. cit.*, pp. 53-62. (This essay was initially presented in 1974); Philippe C. Schmitter, "Still the century of corporatism?", *Review of Politics*, vol. 36, No 1 (Jan. 1974), pp. 85-131; R. E. Pahl & J. T. Winkler, "The coming corporatism", *New Society*, (10 October 1974).

[10] Leo Panitch, "Recent theorizations of corporatism: reflections on a growth industry", *British Journal of Sociology* 31 (June 1980), pp. 159-87.

[11] Jack T. Winkler, "Corporatism", *European Journal of Sociology*, Vol. 17, No 1 (1976), pp. 100-136.

[12] Salvador Giner & Manuel Pérez Yruela, *La Sociedad Corporativa*, (Madrid: Centro de Investigaciones Sociologicas, 1979).

[13] See, for example, the various essays in Klaus Armigeon et al., *Neokorporatistische Politik in Westeuropa*, (Konstanz: Universität Konstanz, Sozialwissenschaftliche Fakultät, Diskussionsbeitrag No 1, 1983).

[14] Cf. Kenneth E. Boulding, *The Organizational Revolution* (Chicago: Quadrangle Books, 1968), originally published in 1953; Samuel P. Hays, *Response to Industrialism: 1885-1914*, (Chicago: The University of Chicago Press, 1957), especially pp. 48-70; Robert Prestus, *The Organizational Society*, (New York: Vintage Books, 1962); Carl G. Gustavson, *The Institutional Drive*, (Columbus, Ohio: Ohio University Press, 1976); Jerry Israel (ed.), *Building the Organizational Society*, (New York: The Free Press, 1972).

[15] Cf. Philippe C. Schmitter, "Interest intermediation and regime governability in contemporary Western Europe and North America", in S. Berger (ed.), *Organizing Interests in Western Europe*, (Cambridge, England: Cambridge University Press, 1981), pp. 287-330, for a criticism of this literature.

[16] Alexis de Tocqueville, *De la Démocratie en Amérique*, (Paris: Gallimard, 1961), Vol. II, p. 117.

[17] Robert Michels, *Political Parties*, (Glencoe: The Free Press, 1949).
[18] Albert Meister, *Vers une sociologie des associations*, (Paris: Editions Economie et Humanisme, (1972), p. 42.
[19] S. M. Lipset, M. Trow and J. S. Coleman, *Union Democracy*, (Glencoe: The Free Press, 1956) in the *locus classicus* for this point.
[20] Mancur Olson, *The Logic of Collective Action*, (Cambridge, Mass.: Harvard University Press, 1965).
[21] Claus Offe & Helmut Wiesenthal, "Two logics of collective action: theoretical notes on social class and organizational from", *Political Power and Social Theory*, Vol. 1 (1980), pp. 67-115.
[22] James Q. Wilson, *Political Organization*, (New York: Basic Books, 1973), pp. 30-91.
[23] Albert O. Hirschman, *Exit, Voice and Loyalty*, (Cambridge, Mass.: Harvard University Press, 1970).
[24] Philip Selznick, *Law, Society and Industrial Justice*, (New York: Russell Sage Foundation, 1969), pp. 37ff. for the first cited item in each of the above paired quotes.
[25] Cf. Frederic S. Burin & Kirt L. Schell (eds.), *Politics, Law, and Social Change: Selected Essays of Otto Kirchenheimer*, (New York: Columbia University Press, 1969) and Stein Rokkan, "Numerical democracy and corporate pluralism", in R. Dahl (ed.) *Political Oppositions in Western Democracies*, (New Haven: Yale University Press, 1966), pp. 70-115.
[26] Peter Bachrach & Morton S. Baratz, *Power & Poverty. Theory & Practice*, (New York: Oxford University Press, 1970), esp. pp. 3-66.
[27] Mancur Olson, in his recent, *The Rise and Decline of Nations. Economic Growth, Stagflation and Social Rigidities*, (New Haven: Yale University Press, 1982), attempts to elevate this contingent point into a major explanatory factor.
[28] Norberto Bobbio, *op. cit.*, p. 29.
[29] Philippe C. Schmitter, *op. cit.*, in S. Berger (ed.), *Organizing Interests in Western Europe.*
[30] Charles Sabel, "The internal politics of trade unions", in *ibid.*, pp. 209-248.

# II

# Law

# Repudiating Montesquieu? The Expansion and Legitimacy of "Constitutional Justice"*

MAURO CAPPELLETTI
Florence/Stanford

## 1. The Contribution of Legal Justice to the Pursuit of Political Freedom

What "human justice" can do is solve, or attempt to solve, concrete problems of individual and societal life: to enact and enforce norms, to create institutions, to design processes, all with that one goal in mind – to solve actual problems. But human problems continuously change; and so do norms, processes, and institutions. Human justice is changing justice – whether or not there is, at a final point, an all-encompassing permanence, an Absolute which gives pause, and meaning, and light to all this moving and striving and passing which is human life.

I am honored to join you today in paying a tribute of respect and gratitude to Pope John XXIII. His vision of, and faith in, the Absolute did not diminish his deep commitment to, and compassion for, the changing. Human justice never ceased to be his concern. His two major Encyclicas, *Mater et Magistra* (1961) and *Pacem in Terris* (1963),[1] represent a formidable effort to trace the guidelines for the solution of, perhaps, the most challenging life-problems of our epoch: the challenges of oppression, of poverty, and of war; the problems of human freedom and dignity, of social justice, and of peaceful coexistence of individuals and peoples – problems whose solution should unite, as he said, all men of good will, whatever their race, nationality, and faith. He thus gave us a basis for designing a philosophy of life for individuals and nations of our epoch.

His was, one might say, at its roots a doctrine relevant for all times, because every epoch has suffered from oppression, social injustice, violence, and war. But in a deeper sense, his doctrine was meant as a philosophy most essentially of and for our time. It was the lesson drawn from the tremendous challenges of our century, for our century is one that,

while uttering the proclamations of the noblest ideals of individual freedom and human dignity, yet has been justly characterized as the epoch of the most terrifying systems of oppression of individuals, groups, and peoples, with the holocausts and the attempted genocide of entire populations and nations.

Ours is also a century that, while witnessing the most impressive growth of material wealth and a unique possibility to create new wealth and welfare, yet has been, and is, plagued by widespread misery and massive starvation. And, our century is one in which a shrinking world demands, and indeed cries out for unity and peace, and yet has been the era of two most dreadful world wars and of the impending menace of a third and ultimate conflict, the last universal deluge, the end, possibly, of humankind.

It is time, is it not, for all of us to be concerned with fundamentals. If such relatively limited episodes as, say, the Watergate scandal, have reminded American lawyers and law teachers that there is a problem of morality in law, in the legal profession, and in legal education; and if Vietnam, but also Iran, and Chile, and Afghanistan, and many more, have reminded men and women – hopefully behind the iron curtain as well – that there is a problem of morality in politics and in the way our endangered world is run; so it is time, is it not, for all of us to be concerned about finding a proper solution to the most fundamental issues – the survival problems of our epoch. The problems of pursuing freedom, justice, and peace are new at least insofar as they have assumed an unprecedented dimension in our time: a dimension that, if adequate solutions are not found, might eventually signify the end of millennia of civilization.

Let me try, then, to use this distinguished forum to make a brief inquiry into a specific answer that our epoch has been trying to give to these major problems of survival, and especially to one of them – the problem of political oppression.

On a personal level, let me add that it is especially rewarding for me to be speaking to you on this topic today, the 25th of April, Italy's *liberation day*, the fortieth anniversary of the liberation of my country from a dreadful system of political oppression which had led to the most tragic of all wars. My inquiry will focus on, without being limited to, Europe; and it will obviously be the inquiry of a lawyer, for I will try to sort out some of the most significant *legal* norms, institutions, and processes that characterize, in my opinion, actual or potential responses to a most dramatic challenge of our time. Others, of course, might emphasize different answers to this challenge – or, as I would prefer to think, different facets of the same answer – in terms of, for instance, economic approaches rather than the legal ones with which I am more directly concerned here and now.

## 2. The Meaning of "Constitutional Justice"

The principal answer in terms of legal justice to the problem of oppression can be expressed with a formula widely used today in Europe: *constitutional justice*. It is used to indicate that governmental power is limited by a constitutional norm, and that procedures have been designed and institutions created to enforce such limitation.

True, the forms of oppression that have characterized our epoch are many and very complex. For instance, nongovernmental power, such as the power of organized groups, economic corporations, unions, callings, and political parties, has at times proved to be no less dangerous and oppressive, no less intrusive into the privacy and freedom of individuals, than the official power of the state. Even the fantastic, splendid developments in technology are themselves a potential menace, for instruments of intrusion have become more and more readily available for oppressive use. And indeed, never, perhaps, as deeply as in our time has the individual felt the oppression of his or her "solitude within the crowd"; the sentiment that our voice is like the Biblical one, *clamans in deserto*; the sense of alienation, which is one of the basic psychological diseases of modern man.

Still, the danger which has proved to be most fearful in our century is, no doubt, that of the organized official power – the state and its manifold branches and agencies, its central and local proliferations. The cases, to mention only the most clamorous ones, of Nazi Germany, of Fascist Italy, of Stalin's Russia shall not have passed without teaching us their important lesson: when the political power is unchecked, then even the instruments of new technology, of mass communication, of so-called "popular education," all can be perverted into a grand corrupting machine. The corruption of the minds is obtained through massive misinformation and the interdiction of all criticism. Remember one of the infamous laws discussed years ago in the much cited debate on law and morals between Professors Hart and Fuller,[2] namely the German law of 1944 that had allowed for a man, denounced by his wife, to be sentenced to capital punishment, his "crime" having been that while at home on leave from the army, he had made critical remarks about Hitler.[3] What can emerge from unchecked government, as sad recent history tells us, is a distortion of even the most common sense of justice – hence intolerance and fanaticism, and, eventually, the acceptance, even the call for, violence and war.

Constitutional justice, I believe, is one and indeed a most important and promising answer that a growing number of nations have tried to give to this problem of governmental oppression. As already mentioned, what is implied by constitutional justice is the adoption of a new[4] kind of *constitutional norms*, *institutions*, and *processes* in an attempt to thereby limit and control the political power. There are, of course, a variety of ways to help achieve that end. These include regionalism, which brings

about a decentralization of at least part of the political power, one form of a "vertical sharing" of that power. Here, however, I intend to center my discussion on judicial review of the constitutionality of state action, and particularly of legislation. This is a development which, in a most real sense, has changed the governmental structure of much of Continental Europe over the last forty years or so, with expansions into other parts of the world, including, for instance, Japan.

## 3. The Rise and Growth of Constitutional Justice in the Post-World War II Era

Austria since 1945, Japan since 1947, Italy since 1948, Germany since 1949: emerging from the nightmare of tyranny and war, all these countries have followed a similar path in their effort to build a new form of government, civil-libertarian and democratic. Each of them has adopted a written Constitution, declared to be binding for all branches of government; they have introduced severe limitations to the amendment process of the Constitution, thus shielding the new basic law from the whims of passing majorities; they have included a bill of rights in the constitution, thus extending the constitution's protection to the individual in his or her relationship with the governmental power; and, last but not least, they have entrusted the enforcement of the constitution, and its bill of rights, to new or renewed judicial tribunals, endowed with important guarantees of independence vis-à-vis the political branches.[5]

This, of course, might seem "old hat" to Americans. Let me suggest, however, that even in the United States the role of constitutional adjudication has acquired its current importance only in the post-World War II epoch, when it became the foremost instrument for the enforcement of certain basic civil rights of individuals and minority groups against reluctant majorities in the states, and against the inaction of the political branches at the federal level. As for the rest of the world, it should be noted that in many other countries constitutional justice, in all its meanings mentioned a moment ago, has represented a fundamental innovation. Indeed, it has been a real revolution at least in Continental Europe and, perhaps, in Japan.

Constitutions and bills of rights, of course, have existed in France, Germany, and elsewhere for many years. Until the post-World War II epoch, however, their meaning tended to be that of political-philosophical declarations rather than that of legally binding enactments. For, with few sporadic and short-lived exceptions (most notably that of Austria in the 1920's and early 1930's[6]), no independent body was entitled to supervise their actual application. The constitutional revolution – and I do mean what this word says – occurred in Europe only with the suffered acquisi-

tion of the awareness that a constitution, and a constitutional bill of rights, need judicial machinery to be made effective. The United States certainly provided an influential precedent. But the most compelling lesson came from domestic experience, the experience of tyranny and oppression by a political power unchecked by machinery both *accessible* to the victims of governmental abuse, and *capable* of restraining such abuse.

The lesson was eventually learned: constitutional courts have been created, and constitutional processes have been designed to make them work. Let me mention only one such process, for it seems most indicative of the philosophy permeating this constitutional and civil rights revolution. In Germany, in 1951 ordinary legislation gave standing to any individual to bring a complaint before the newly instituted Constitutional Court against any kind of state action, legislative, administrative, or judicial that violated his or her constitutionally entrenched rights.[7] In 1969 this extraordinary remedy, called *Verfassungsbeschwerde* or constitutional complaint, was made part of the German Constitution; and Austria, especially since 1975, has adopted a similar process.[8] Through this and other devices, the constitutionality of thousands of legislative and other governmental actions has been controlled and the people's fundamental rights have been protected by independent courts in Germany, Austria, Italy, and elsewhere.

The success of "constitutional justice" as an instrument for the protection of human rights, and its profound impact on the democratic-libertarian form of government, have been generally recognized in all those countries, even though, naturally enough, dissent is often voiced about the contents of particular constitutional decisions or even about some general trends in the constitutional case law. The most conclusive evidence, perhaps, of the success of this phenomenal development is offered by its tremendous force of expansion. Let me mention just a few episodes: Cyprus in 1960, Turkey in 1961, and Malta in 1964 all have introduced forms of constitutional adjudication largely modeled after those of Germany, Austria, and Italy.[9] Indeed, it seems as though no country in Europe, emerging from some form of undemocratic regime or serious domestic strife, could find a better answer to the exigency of reacting against, and possibly preventing the return of, past evils, than to introduce constitutional justice into its new system of government. This has been the case for Greece in 1975, after the fall of the regime of the colonels;[10] for Portugal in 1976, after the fall of the Salazar regime;[11] and for Spain in 1978, after the fall of Franco.[12] Significantly, Yugoslavia too, in its quest for political and ideological autonomy vis-à-vis the Soviet Union, enacted a constitution in 1963 that introduced a system of judicial review.[13] Yugoslavia has been the first and, so far, the only country under a Communist regime to do so; but it is most meaningful that Czechoslovakia in 1968 – the year of the passions and hopes of the "Spring of Prague" – tried to

follow suit,[14] and that so did Poland in 1982 before "Solidarnosh" and all the rest were condemned to silence.[15] Unlike Yugoslavia, however, the constitutional amendments of both Czechoslovakia and Poland have remained dead letter, crushed by the resurgence of their autocratic regimes. Indeed, if one wisdom clearly emerges from comparative analysis of these most recent developments, a wisdom that many critics of the democratic legitimacy of judicial review seem to neglect, it is that no effective system of judicial control is compatible with, and tolerated by, anti-libertarian, autocratic regimes, whether they place themselves at the left or the right of the political spectrum. This fact, that judicial review is anathema to the tyrant, is confirmed by developments in many countries in several continents, and most frequently in Latin America and Africa.[16] A peculiar illustration is offered by South Africa where a "constitutional crisis" developed in the 1950's when the judiciary declared certain discriminatory enactments of the South African Parliament to be unconstitutional. The crisis culminated with the adoption of South Africa's Constitution of 1961 which effectively denied the judiciary any authority to review the validity of legislation.[17]

## 4. Has France Repudiated Montesquieu?

I am not an expert on Japanese affairs. Yet I am told by those who know that even in that nation, constitutional justice, initially seen by many as an alien element in the Japanese system of government, has gradually carved out for itself a relevant place and a genuine significance within that system, even though not, or not yet, as great a place and significance as in Continental Europe.[18]

Returning to Europe, my story would be much too incomplete if I did not elaborate somewhat on those two other great nations, France and England. These countries have been, and in part still are, much more reluctant than most of the rest of Europe to participate in the "constitutional revolution." Parliamentary supremacy has long been their political credo – the national Parliament, as the embodiment of the democratic will, being thought to be immune from judicial control. This has been the tradition, and the myth, both of England since the "Glorious Revolution" of 1688, and of France since its Revolution one century later, a myth not shared by the American Revolution.[19]

To be sure, each of those two European nations has a different history of Parliamentary supremacy. In France it must be traced back, in part, to a deeply-felt popular revulsion against the abuse of the judicial office by the higher courts of justice under the *ancien régime*. These courts, whose name, ironically, was *Parlements*, asserted their power to review acts of the sovereign, refusing to apply those found to be incompatible with the

"fundamental laws of the realm."[20] The reading by those courts of such – mostly unwritten – fundamental laws, however, led the courts to affirm the *heureuse impuissance* of the legislator to introduce even minor liberal reforms. Those judges were so deeply rooted in the feudal regime that they found any liberal innovation unacceptable. Their office was hereditary and could be bought and sold,[21] and their activity was to be paid by each individual litigator as though it were their privilege, not their duty, to administer justice.[22] Their status, education, and personal, family, and class interests combined to motivate their extremely conservative attitude, an attitude which eventually contributed to the triggering of the revolutionary explosion.[23] The popular feelings against the *Parlements* were well justified,[24] and this justification is reflected, albeit in a veiled form, in that celebrated work, *De l'esprit des lois*, first published in 1748 by one who, when speaking of the judges of his time and country, knew only too well what he was saying. Charles-Louis de Secondat, the offspring of an ancient family of judges "parlementaires," at the age of twenty-seven, in 1716, had already become "Président à mortier" at the *Parlement* of Bordeaux, having inherited the high judicial office, as well as the name of Montesquieu, from his deceased uncle.[25] Quite understandably given the kind of judges of the time, an enlightened Montesquieu preached that the judges should be entrusted with no political power at all:

> There is no liberty . . . if the power to adjudicate is not separated from the legislative and the executive powers.[26]

Even if the law, "which," he said, "is at the same time clairvoyant and blind,"[27] should appear in certain cases to be too harsh, still it is not for the judges but only for the legislator to intervene. To the judges appertains only the duty to blindly apply the law, for

> the judges of the nation are . . . nothing but the mouth which pronounces the words of the law; they are inanimate beings who cannot moderate either the force or the rigor of the law.[28]

Thus, although Montesquieu, unlike Locke, did list the judiciary as one of the "three powers," coming after the legislative and the executive powers,[29] he also made it clear, however, that the third branch, in a real way, is no "power" at all:

> Of the three powers of which we have spoken, the judicial is, in a sense, null.[30]

Whatever the actual influence of Montesquieu on the French Revolution, this idea was to become a central part of its ideology. The Revolution proclaimed as one of its first principles the absolute supremacy of statutory law, the law enacted by the *corps législatif* as the representatives of the

people, while demoting the judiciary to what was seen as the purely mechanical task of applying that law to concrete cases. Of course, also the Rousseauian faith in the "infallibilité" of the "loi" as the expression of the "volonté générale" found its triumph in this Revolutionary development.[31] To be sure, the strict separation, "French style," of the governmental powers, whether or not actually "Montesquieuian" in inspiration,[32] was miles away from the kind of separation of powers which almost contemporaneously was adopted by the American Constitution. Separation of powers in America is better described as "checks and balances";[33] under this principle, an extremely important role of review of both administrative and legislative action was to be reserved to the courts. *Séparation des pouvoirs* French style, on the contrary, implied that the judiciary should assume a role totally subservient to, and at any rate strictly separate from, the role and activity of the political branches; as such, it soon proved to be the source of problems and difficulties no less serious than those it was intended to solve. The legal history of France throughout most of the 19th century is a continuous illustration of such problems, as well as of the striving efforts to find new and more appropriate solutions for these problems. Reducing the judicial function to a blind, "inanimate," slot-machine application of the laws to individual cases is oblivious of the reality, that is, of the fact that no norm, law, or code can be so clear and complete as to allow for only one "correct" interpretation.[34] More importantly still, the Montesquieuian (and Rousseauian) approach, as implemented by French Revolutionary legislation, while intended to protect against tyranny, left the doors wide open to both legislative and executive tyranny. The famous Revolutionary *loi* of 16-24 August 1790 on "organisation judiciaire," whose principles were to become the pillars of the French judicial system and other Continental systems influenced by the French, established that no control whatsoever by the judiciary was allowed either of legislative or of administrative action:

> *Title II, Art. 10*: The judicial tribunals shall not take part, either directly or indirectly, in the exercise of the legislative power, nor impede or suspend the execution of the enactments of the legislative body. . .
> *Title II, Art. 12*: [The judicial tribunals] shall refer to the legislative body whenever they find it necessary either to have a statute interpreted or to have a new statute.
> *Title II, Art. 13*: Judicial functions are distinct and shall always remain separate from administrative functions. Under penalty of forfeiture of their offices, the judges shall not interfere in any way whatsoever with the operation of the public administration, nor shall they call administrators to account before them in respect of the exercise of their administrative functions. . .[35]

This meant that both legislators and public administrators were exempt from any check by a third, independent, nonpolitical or, at least, less political branch. *Internal* controls, of course, could be and, in fact, were

designed. But history, and often unhappy history at that, has proven that, to be effective, controls of the political branches can hardly be controls *from within*. Efficient executive power is hierarchical; at its top level it does not easily allow for independent internal control; and this is no less true for a legislative power which affirms itself as supreme. It should come as no surprise, then, that all systems past and present of political, nonjudicial control, such as those experimented with in France under the Constitutions of 1799, 1852 and 1946, and those currently adopted by most communist countries have proved to be utterly inefficient.[36] Effective control of the political branches must be control *from outside*: it must be entrusted to persons and agencies sufficiently independent from those controlled. And this, in fact, is what the French have gradually realized, at least as far as administrative action is concerned. A glorious institution, the *Conseil d'Etat*, gradually evolved in the 19th century from what initially was a mere department internal to the administration, into an independent judicial agency fully recognized as a high court of France. Its role is to control the legitimacy of administrative action. The more the *Conseil*'s role was to become important and accepted, the more independent the *Conseil* was to become vis-à-vis the political branches. And, with its independence, the judicial nature of its process also became more and more pronounced and recognized, with all the consequences of such development – including the adoption of those safeguards most characteristic of the judicial process: the impartiality of the adjudicator, the right of the parties to be heard, and all the many corollaries of these basic rules of "natural justice."[37]

France, of course, was the initiator, on the Old Continent, of this development, the establishment of *justice administrative* or judicial review of administrative action. Sooner or later, however, other Continental countries followed its lead, and so the French system of *justice administrative* became the model for the development of its analogues, *Verwaltungsgerichtsbarkeit* in Germany, *giustizia amministrativa* in Italy, etc.[38]

Our century, however, was to teach us yet another lesson: that the Rousseauian idea of the infallibility of Parliamentary law is but another illusion,[39] for even the legislative, not only the administrative branch might abuse its power; that this possibility of legislative abuse has grown tremendously with the historical growth of legislation in the modern state;[40] also, that legislatures might be made subservient to uncontrolled political power, and that legislative and majoritarian tyrannies can be no less oppressive than executive tyranny. Suffice it to remember Fascist legislation depriving the Jews and other minorities of their most basic rights. This is why Austria and Italy and Germany, surfacing from the moral and material ruins of political perversion, dictatorship and defeat, soon turned to *constitutional justice*, as previously explained, to create a new kind of control to be added to the more traditional one of *administra-*

*tive justice*. In so doing, they attempted to limit and check the power of the legislature and legislative majorities, within the scheme of the new constitutional norm made binding by constitutional adjudication. The historical influence of French ideas, however, can also help us to understand why in order to do so all these countries felt it necessary to follow a path similar to that of *justice administrative*: they all created a new type of organ of control – almost as a *pendant* of the 19th-century *Conseil d'Etat* and its German and Italian analogues, the judicial organs of control of administrative action.[41]

France, on the other hand, was somewhat less immediately involved in this newer course of action. The abuses of the Vichy *régime* during World War II had been less outrageous, perhaps, and certainly less durable than those in the other countries. This might explain why France, although the leader in the 19th-century development of *justice administrative*, has not been a leader in the post-World War II development of "constitutional justice."

This is not yet the end of the French story, however. If not the leader, France, too, has become involved lately, and quite deeply so, in this newer development, the constitutional and judicial review revolution.[42] This has become apparent especially since 1971, when a body created by de Gaulle's Constitution of 1958, called the *Conseil Constitutionnel*, most courageously transformed its role, and transformed it radically. Originally envisaged as a mere watchdog of the enlarged powers of the executive under the *Général's* régime, in July 1971 the *Conseil Constitutionnel* asserted itself for the first time as an independent, quasi-judicial organ whose role is to review the constitutionality of Parliamentary legislation violative of fundamental rights. A constitutional amendment of 1974 under President Giscard d'Estaing reinforced this development by granting Parliamentary minorities standing to challenge legislation before the *Conseil Constitutionnel*. Today many an expert would agree with our French colleague, Doyen Louis Favoreu, when he maintains that the system of judicial review of the constitutionality of legislation developed in France during the last fifteen years or so, is almost as effective as those of neighboring Continental nations.[43] At least two serious limitations of the French system should not be neglected, however.[44] First, there is no possibility in France for individuals whose fundamental rights have been violated to bring their complaint to the *Conseil Constitutionnel*; Parliamentary legislation can be attacked only by a minority of at least 60 members of either Chamber of Parliament, or by a handful of political authorities having individual standing to do so in the general interest. Second, legislation can be attacked only during the short period between its enactment by Parliament and its promulgation; once promulgated, no judge in France can set aside a *loi* by declaring its conflict with the Constitution. And yet, even with these limits, it took only a few years for

judicial review of legislation in France to gain a remarkable importance. In many cases constitutional rights of minorities and individuals have found in this review system a formidable shield against what was felt by many as majoritarian abuse. Thus the French constitution, and most particularly its bill of rights which includes by reference the 1789 *Déclaration des droits de l'homme et du citoyen*, has for the first time become in a full sense a legally binding, judicially enforceable document.[45]

## 5. England's "Grundnorm": The Absolute Supremacy of Parliament

England, of course, presents us with a much different story. On the one hand, contrary to *ancien régime* France, there have been no deeply felt popular feelings in England against the judiciary, whose historical role in protecting individual liberties has generally enjoyed widespread respect.[46] This can explain why, unlike France, judicial review of *administrative* action has never encountered serious obstacles in Great Britain. The doctrine of separation of powers was never fully adopted in England in its "French version," that is in the version which implied the prohibiton of any "interference" by the courts even with the administrative, not only with the legislative, branch. On the other hand, the English Revolution of 1688 did affirm, and very strongly so, the absolute supremacy of Parliament which, as the proverb goes, "can do anything except transform a man into a woman or a woman into a man."[47] Rejecting such judicial precedents as the famous decision by Lord Coke in *Dr. Bonham's Case* of 1610,[48] Parliamentary supremacy had as its corollary the unreviewability of Parliamentary legislation – the "omnipotence" of positive (statutory) law and the judicial powerlessness to control the "validity" of that law.[49]

If the French brand of judicial powerlessness might find in Montesquieu its most authoritative, though not unambiguous, theorist, the great liberal thinker and theorizer of the Glorious Revolution, John Locke, might be seen to have played a similar role in England. Although frequently associated with the historic doctrine of separation of powers, Locke in fact did not even view the judiciary as a separate "branch" or "power." In his trichotomy, the two "derivative" or "inferior" powers are the "executive"[50] and the "federative,"[51] whereas the "supreme" power, the "legislative,"[52] is magnified as "the Soul that gives Form, Life, and Unity to the Commonwealth."[53] Even though Locke's "legislative" was bound to "pronounce" and "enforce" the "eternal and immutable laws of nature," which are found, not created, by reason,[54] he did not see the judiciary as the authorized and privileged enforcer of these natural law limits of the legislative will.[55] Locke's doctrine was to be echoed and made more explicit by Blackstone when the great commentator bluntly rejected judi-

cial review as being tantamount to setting "the judicial power above that of the legislature, which would be subversive of all government."[56]

Unlike France, this is not past history for England. Parliamentary supremacy is still affirmed as the basic principle, the *Grundnorm*[57] of the unwritten constitution of that country. Yet, significant breaches have been opened over the last few years into the principle's solid, tricentenary walls. I shall mention two of them, which apply to the United Kingdom but at the same time to much of the rest of Western Europe as well. For they lead us to a new and most unique dimension in the extraordinary development and growth of judicial review in Europe – its *transnational dimension.*

## 6. Is England Abandoning Her Lockeian Grundnorm? Community Law "Cannot Be Held Back"

A first "breach in the walls" has been opened by Community law. As you know, since 1973 the United Kingdom has become a full member of the European Community – the so-called Common Market, in which now ten countries of Europe participate, soon to become twelve countries with a population of over 300 million people. One of the basic features of the Community is that it has been entrusted with law-making powers in a wide variety of areas, especially of economic, but also of social concern. Community law, mostly enacted by the Council of Ministers of the European Community with some participation of the Community Commission and the European Parliament, has indeed proved to be an expanding body of transnational legislation, primarily consisting of, by now, thousands of so-called "regulations" and "directives."[58] In the colorful expression of Lord Denning, this body of Community legislation penetrates the British legal system, and the systems of the other nine member states as well, "like an incoming tide. It flows into the estuaries and up the rivers. It cannot be held back."[59] And the reason why it cannot be held back is that a basic principle of Community law affirms the "direct applicability" of Community law as if it were, automatically, *the law of the land* of each of the member states.[60] Again in Lord Denning's words, "Parliament has decreed that (Community law) is . . . to be part of our law."[61]

It belongs to the very nature of Community law, moreover, that, as a general rule at least, it must be uniform and uniformly applied in all the member states. This explains why, since at least 1964, a consistent stream of decisions of the European Court of Justice – the Court of the Community sitting at Luxembourg – has established that Community law not only is the law of all the member states, to be directly applied by all the national courts, but that it is, moreover, the *higher* law of the member states, prevailing over conflicting national legislation.[62] No matter when

enacted, national legislation must be set aside by all the national courts of the ten countries if they find it to be contrary to Community legislation;[63] and problems of interpretation are decided, in last resort and with final effect for all the member states, by the European Court of Justice.[64]

Here we can see, then, that an important new form of judicial review of legislation has penetrated into "the estuaries and the rivers" of the British as well as the other nine European systems. It is a form of review quite akin to the American adjudicatory enforcement of the supremacy of federal legislation over conflicting state law. To be sure, it is not a control of the *constitutionality* of legislation; yet it is at least a first important move toward recognizing, even in England, that the historical principle – the absolute supremacy of Parliamentary law – does no longer fully prevail.

## 7. The "Transnational Justice" of the European Court of Human Rights

A second development is even more akin to our concern, that is, to review of the *constitutionality* of legislation, and most particularly to judicial review as an instrument to protect human rights. Indeed, a few years ago the development I am now going to discuss motivated the distinguished American constitutionalist Charles Black – a past "Pope John XXIII lecturer" as I have learned – to maintain that England, contrary to generally accepted opinion, *does* already have a written and binding bill of rights.[65]

This second episode materialized especially when the United Kingdom, a ratifying member of the European Convention of Human Rights since 1951, in 1966 accepted the optional clause of Article 25 of that Convention.[66] This clause creates a form of transnational *Verfassungsbeschwerde*; after exhaustion of the national remedies, it grants standing to all individuals to bring before the Convention's judicial machinery in Strasbourg their complaints against any sort of state action, including legislation, violative of their rights entrenched in the Convention. Let us be reminded that the Convention is a comprehensive transnational bill of rights; with only one exception, Finland, it is now adhered to by all the countries of Western Europe, twenty-one countries encompassing a population of more than 350 million people.[67]

Thus a transnational bill of rights has become binding for England, among other nations, and is enforceable by a transnational adjudicator to whom the British citizens have access. From this it is perhaps only a small step to accept the idea that the Convention is part of the law of England, and indeed binding also for the British Parliament; and that the British courts must enforce such higher law, lest their judgments be subject to the

condemnation of the transnational adjudicators at Strasbourg.[68] Whether the Britons are ready to take this further step and thus adopt a full-fledged system of judicial review is not a question I want to discuss now. Suffice it to say that complaints successfully brought under the Convention by individuals against British legislation and other British state action have been rather frequent in the last several years, and in quite a few cases the condemnations by the European Commission and Court of Human Rights have aroused bad feelings in the United Kingdom, for they have cut into cherished traditions; nevertheless, British authorities, including the Parliament, have generally shown their willingness to comply with the final decisions of the European Court of Human Rights. *De facto* at least, the supremacy of the transnational bill of rights has been largely confirmed in Western Europe.[69] Great Britain , in particular, while ostensibly sticking to its tradition of rejecting judicial review of legislation, has gone a long way by now toward repudiating its own brand – a "Lockeian" and "Blackstonian" brand, we might be tempted to say – of the doctrine of the judicially uncontrollable supremacy of the legislative will.[70]

## 8. On the "Mighty Problem" of the Democratic Legitimacy of Constitutional Justice

We have seen how judicial review has been recently introduced, and/or has greatly expanded its role, in a large number of nations. Indeed, to be complete our list should have been extended to many more countries, including Sweden, especially since 1980,[71] and Canada, especially since its new constitutional "Charter of Rights and Freedoms" of 1982.[72] I should have mentioned, moreover, that even in its most striking, historically unprecedented dimension – the transnational dimension – the European precedent is no longer alone. The American Convention on Human Rights signed in San José of Costa Rica in 1969 has become binding for several Latin-American countries since 1978; largely modeled after the European Convention, this transnational bill of rights has led to the creation, in 1979, of an Inter-American Court of Human Rights sitting in San José and open, perhaps, to future developments similar to those, already quite sensational, of its European antecedent.[73]

But my time is almost over, and I have still to discuss, although briefly, the basic question about the significance and legitimacy of judicial review, in light of its tremendous growth in the contemporary world.[74]

For many nations, as we have seen, this significance has been primarily one of a reaction against past governmental abuses. This has been most evident in several of the countries we have mentioned; and others could be added, even from Africa, Asia, and Latin America. In these continents, too – and especially in Latin America, where some aspects of the phenom-

enon of judicial review are even older than in Europe[75] – comparative study has proved that judicial review of the constitutionality of legislation and other state action has, at least, the potential to work as an instrument to protect individuals and minorities. This is true even though the efficacy of judicial review in the developing world has been much too frequently jeopardized both by insufficient judicial independence and by the use and abuse on the part of the executive of the power to suspend constitutional guarantees.[76] But even in countries, such as England, where, happily, there has been no such legacy of serious governmental abuse, judicial review has been emerging, indirectly at least, as an element of a new and most fascinating trend in law, politics, and human rights: transnationalism. Vertical sharing of power and the ensuing pluralism of legal sources, a typical product of transnationalism as well as federalism, brings about, inevitably, the possibility of conflicts among the various levels of power, of laws, and of rights; and judicial review is the natural instrument to settle such conflicts.

There is, moreover, a general perception at least in Western countries, that in our "age of statutes" – as Dean Calabresi has called it[77] – the control by an independent adjudicator of a more and more pervasive legislator, whose role in the modern state has grown to unprecedented dimensions, is a valuable safeguard; indeed that such control is the necessary "crowning" of the rule of law. True, the legislator in democratic societies is the representative of, and accountable to, the people, whereas it belongs to the very nature of the judicial function that judges shall not be so accountable. The paradox – entrusting unaccountable judges with the function of controlling accountable politicians – is merely apparent, however. In our societies judges are nonaccountable only in the sense that they are not and shall not be held responsible to the other branches or to the people for their individual decisions and philosophies. Their nonaccountability, however, holds only in the short and medium term. There are many ties which, in the long term at least, connect them with their time and society. These ties might be reinforced by the manner of the judicial appointment, as in the United States,[78] or, as in Europe, by the fact that the judges' tenure in office, which surely must be long enough to give them autonomy and assurance, is limited to a given number of years with, as a rule, no possibility of being extended. It should also be noted that the very nature of the judicial process is a highly participatory one, for the judges' role is based on real life cases and can be exercised only upon, and within the limits of, the interested parties' complaints and demands. In this sense, there is at least a high potential for a continuous contact of the judiciary with society's real problems, needs, and aspirations.[79] A sound product of our freedom of speech, moreover, is the fact that judges, too, day after day are subject to public criticism.[80] When we speak of separation of powers today, we certainly do not mean *séparation* in the original

French significance; we mean, rather, reciprocal connections and mutual controls. The judicial nonaccountability is a *political* and a *legal* nonaccountability – and even that with important limitations in case of serious abuses; it is not, however, a *societal* nonaccountability.[81] Abuses of a kind analogous to those of the judges of *ancien régime* France would be hardly conceivable in our societies, for those were the abuses of a *corps séparé,* a group totally separated from the rest of society.

The "mighty problem" of the legitimacy of judicial review cannot be solved by means of purely speculative, abstract solutions valid for any place and time. Indeed there are no such universal solutions; and surely a page of realistic comparative analysis can be more worthy than many books of such abstract speculations.[82] Should our judges today be of the kind that prevailed in pre-Revolutionary France, then of course judicial review would be hardly legitimate. But in our Western world, in which the roles of the political branches have grown into so many areas of our life, and indeed inevitably so, the scrutiny of a more "detached," though not literally "separate," judiciary can be most salutary. Values which are more enduring can be better preserved;[83] individuals and groups that would be otherwise emarginated or oppressed can be better protected; and, more generally, the fairness and the permanent representativeness of the political process itself can be better assured.[84] The democratic principle requires that everyone should have a "voice" in the political process and that it be possible for the minority of today to become the majority of tomorrow. If basic rights such as the freedoms of speech, of opinion, of association could be limited, without due process, by the majority of today, the very democratic principle would be impaired; and this is no less true for the "new rights" of a social and economic nature, for their rationale is to make effective the most basic of all democratic entitlements – the right of access to the legal and political system.[85] Thus constitutional justice, far from being inherently antidemocratic and antimajoritarian, emerges as a pivotal instrument for shielding the democratic and majoritarian principles from the risk of corruption. Our democratic ideal, at any rate – let this point be firmly stressed – is not one in which majoritarian will is omnipotent. And our philosophy of life is not one in which everything can be bargained.

## 9. The Contemporary Human Rights Revolution and Its Legitimacy - Overcoming the Traditional Conflict Between Natural and Positive Law

Let me sum up now, by way of conclusion, the two major theses I have tried to convey.

The first is that, since World War II, Western societies have been

living what I do not hesitate to characterize as a constitutional and civil rights revolution. Indeed, at some points there have been signs of this trend going even beyond the Western world: I should only mention the Universal Declaration of Human Rights of 1948 and the International Covenants of 1966 in force since 1976.[86] But unfortunately, these documents have not been accompanied by legal processes and institutions strong enough to give even an initial degree of effectiveness to the rules incorporated in them.[87] And yet even these attempts, as embryonic as they remain, are significant for they witness a universal aspiration, although largely unfulfilled.

Our skepticism, of course, and the many implementation failures and gross violations of the human rights philosophy might often hide this humanistic feature of our time, and indeed many developments might obscure it, even deny and ridicule it. We shall reject unconstructive, excessive skepticism, however, as well as its twin brother, nihilism – these recurring diseases of intellectual narcissism. We are convinced, with the great philosopher of "moderate skepticism," David Hume, that excessive skepticism cannot resist the test of "action" and the reality of "common life."[88] And this reality demonstrates that, as Professor Louis Henkin affirms and amply documents, there has been an "explosion" of human rights in the "libertarian democracies" of our contemporary world.[89] This explosion has been characterized by an unprecedented concern for the creation of effective instruments, national and transnational, if not yet universal, to protect the basic rights of individuals and groups – including the poor, racial and religious minorities, the young and the old, women and, more generally, those traditionally deprived of fair and equal access to the law. Not to recognize the historical importance and the unprecedented character of this ongoing development is to be deaf and blind to, perhaps, the most phenomenal societal transformation trend that has ever occurred in human history.[90]

This is far from a rosy vision of our epoch. Indeed, the human rights explosion is but one attempt to give an answer to those problems which, more than ever before, as I said at the beginning of this talk, are endangering humankind's civilization and survival – the problems of oppression, of poverty, of war. Whether the attempt will fail or succeed is for the future to decide. But it seems clear to me that, if this attempt will be successful, national and transnational human rights and their judicial enforcement will have a good share of the merit. Let me make this statement very clear – that I see no future for humankind, unless a renewed philosophy of tolerance and mutual respect, in a real sense a human rights philosophy, will enable us to make decent use of the tremendous material power we have acquired.

The unprecedented expansion of judicial control of the political branches is a nonsecondary facet of this human rights revolution. The very

fact that, until the post-World War II epoch, judicial review in the United States, while playing an important role in the formation of "a more perfect Union," did not play a comparable role as an instrument for the protection of civil rights,[91] seems to prove my point. For only in our epoch has the time become ripe for what I insist in calling our civil rights revolution. A writer has one of his "characters" say that only danger and suffering make human beings sensitive to justice, to the feelings and the inquiry of what is bad and good – in sum, to human values.[92] The tensions and dangers of our time are so great and imminent, that this sense of values in some way has been and will be forced to emerge – hopefully, not only in the West. And, in our Western societies it has been a privileged, though surely not exclusive, role of constitutional and transnational judges to interpret and sort out those values that cannot be compromised.

My second thesis has been that this judicial role is a legitimate one. Surely we might disagree, even fight against, certain determinations or trends in constitutional adjudication. Still, a century and a half of Continental history is there to demonstrate that the alternative solution is worse indeed. In the absence of judicial control, the political power is more easily exposed to the risk of perversion. Judicial control, of course, is no infallible remedy; as a bulwark of our freedoms, it might often prove to be too weak to resist tyranny, as the experience of many countries demonstrates. If not an invincible bulwark, however, it may at least act as a warning and a restraint.

Does this development mark the revival of a new "natural law"? Many have said so.[93] I would go farther and say that modern constitutionalism, with its basic ingredients – a civil-libertarian bill of rights and judicial enforcement of it – is the only realistic attempt to implement natural law values in our real world. In this sense, our epoch, if any, is the epoch of natural law. More accurately, however, I would say that modern constitutionalism is the attempt to overcome the pluri-millenary contrast between natural and positive law, the contrast, that is, between an immutable, unwritten higher law rooted in nature or reason, and a passing law written by a particular legislator of a given place and time. The modern constitutions, their bills of rights, and judicial review are the elements of a "positive higher law" made binding and enforceable: they represent a synthesis of a sort – a Hegelian synthesis as it were – of legal positivism and natural law. They reflect the most sophisticated attempt ever designed to "positivize" values without, however, either absolutizing such values or relinquishing them to the mutable whims of passing majorities.[94]

Let us condemn judicial decisions that in our perceptions are wrong. But let us be aware that there is a worth and a legitimacy in an institution whose very *raison d'être* is to control the political power and to protect us against abusive exercise of that power. If it is true, as I think comparative study amply demonstrates, that in the post-World War II era judicial

review in very many countries has been a valuable instrument to reinforce our fundamental freedoms, then its democratic legitimacy is also confirmed. For anything which can reinforce the freedom of the citizens, surely reinforces democracy as well.[95]

## Notes

* Copyright © Mauro Cappelletti. This is a slightly revised version of an article that was originally published in 35 *Catholic University Law Review* 1-32 (1985). It is the annotated text of the 20th annual Pope John XXIII Lecture delivered on April 25, 1985, at the Catholic University of America in Washington, D.C. It was prepared within the framework of the project on "Dimensions of Justice: Constitutional and Transnational Enforcement of Human Rights," which the author has directed at the EUI since 1984, and draws upon research carried out at the EUI in two previous projects also directed by the author, namely the "Access to Justice Project" (1976-79) (results published in M. Cappelletti, gen. ed., *Access to Justice*, Vols. I-IV, Milan & Alphen aan den Rijn, Giuffrè & Sijthoff & Noordhoff, 1978-79; and M. Cappelletti, ed., *Access to Justice and the Welfare State*, Alphen aan den Rijn & Brussels, Sijthoff & Bruylant, 1981) and the project on "European Legal Integration in Light of the American Federal Experience" (1979-85) (results published in M. Cappelletti, M. Seccombe & J. Weiler, eds., *Integration Through Law: Europe and the American Federal Experience*, Vols. I (3 bks.)-VI, Berlin & New York, Walter de Gruyter (1985).

[1] An English version and commentary can be found in *The Encyclicas and Other Messages of John XXIII*, Arranged and Edited by the Staff of The Pope Speaks Magazine, Washington, D.C., TPS Press, 1964.

[2] H.L.A. Hart, "Positivism and the Separation of Law and Morals," 71 *Harv. L. Rev.* 593-629 (1958); Lon L. Fuller, "Positivism and Fidelity to Law – A Reply to Professor Hart," *ibidem*, 630-672.

[3] Hart, *op. cit.*, 618-619; Fuller, *op. cit.*, 654-655.

[4] For some historical precedents, however, *see, e.g.*, M. Cappelletti and J.C. Adams, "Judicial Review of Legislation: European Antecedents and Adaptations," 79 *Harv. L. Rev.* 1207-1224 (1966); M. Cappelletti, *Judicial Review in the Contemporary World*, Indianapolis, Bobbs-Merrill, 1971, 25-43 (hereinafter cited as *Judicial Review*).

[5] On these developments *see Judicial Review, supra* note 4, especially ch. 3. The major characteristic of the European systems of review is their "centralized" nature, meaning that the authority to declare a law unconstitutional, hence (in principle) null and void, is reserved to a newly created constitutional court. If faced with the question of the unconstitutionality of a law relevant in the case at hand, the other courts shall not decide that question but rather suspend the case and refer the issue of constitutionality to the constitutional court, whose decisions have *erga omnes* effect. The European systems are contrasted to the American "decentralized" system in which all the courts have the judicial review power. *See Judicial Review, supra* note 4, chs. 3-5.

[6] *See id.*, at 46-47 *et passim.*

[7] *See id.*, at 22-23, 78. To prevent abuse of this process, exhaustion of the normal remedies is required, but even this prerequisite is unnecessary in the case of a complaint reflecting a "general interest" or when delay would cause important prejudice to the party concerned. *See, e.g.*, K. Schlaich, "Procédures et techniques de protection des droits fondamentaux. Tribunal Constitutionnel Fédéral Allemand," in Louis Favoreu (sous la direction de), *Cours constitutionnelles européennes et droits fondamentaux*, Paris, Economica, 1982, 105, at 128-129 (hereinafter cited as *Cours constitutionnelles*).

[8] *See* the report by L. Favoreu on Western Europe, § 8-10, presented to the Conference of the International Association of Legal Science held in June 1984 in Uppsala, Sweden, on "Judicial Review of Legislation and Its Legitimacy – Recent Developments." The regional reports, as well as Mr. Cappelletti's general report are being published in volume form under the editorship of Professors Louis Favoreu and J.A. Jolowicz (hereinafter cited as "Uppsala Reports"). (An adapted version of Dean Favoreu's report is published in *Revue du droit public et de la science politique en France et à l'étranger,* 1984, 1147-1201.)

[9] Constitutions of Cyprus (1960), Turkey (1961) and Malta (1964). *See Judicial Review supra* note 4, at 50-51.

[10] Greek Constitution of 1975. *See* E. Spiliotopoulos, "Judicial Review of Legislative Acts in Greece," 56 *Temple Law Quarterly* 463 (1983); J. Illiopoulos-Strangas, "Grundrechtsschutz in Griechenland," *Jahrbuch des öffentlichen Rechts,* 1983, 396; V. Perifanaki Rotolo, "La Corte Suprema Speciale nella Costituzione greca del 1975," 29 *Rivista trim. di diritto pubblico* 183 (1979); L. Favoreu, *supra* note 8, § 27-31. Under the Greek system, all courts have the power to deny application of unconstitutional laws (so-called "decentralized" system of review, *see* note 5 *supra*), but a newly instituted "Special Supreme Court" has the final word in case of conflicting opinions among higher courts.

[11] Portuguese Constitution of 1976. Especially after the constitutional reform of 1982, Portugal has adopted a system of judicial review similar to that prevailing in the majority of the of the European nations mentioned in the text, entrusting the control function to a newly created constitutional court. *See* L. Favoreu, *supra* note 8, § 22-26; H. Fix Zamudio, *La protección jurídica y procesal de los derechos humanos ante las iurisdicciones nacionales,* Madrid, Civitas, 1982, 203-207.

[12] Constitution of Spain of 1978, followed by the creation in 1980 of a very active constitutional tribunal. *See* E. García de Enterría, *La constitución como norma y el tribunal constitucional,* Madrid, Civitas, 2nd ed. 1982; J. Gonzales Pérez, *Derecho procesal constitucional,* Madrid, Civitas, 1980; M. Aragòn Reyes, "El control de constitucionalidad en la Constitución española de 1978," 7 *Revista de Estudios Políticos* 171 (1979); H. Fix Zamudio, *supra* note 11, 197-202; Favoreu, *supra* note 8, § 13-14.

[13] The system was reaffirmed in the Yugoslav Constitution of 1974. *See* the Uppsala Report by Pavle Nikolić on the Socialist countries, *supra* note 8, § I.1, III.2.A *et passim; see also* H. Fix Zamudio, *supra* note 11, 208-212; *Judicial Review, supra* note 4, 51-52.

[14] *See* Uppsala Report by P. Nikolić *supra* note 13, § III.2.B; Uppsala General Report by M. Cappelletti, *supra* note 8, § 2.

[15] *See* A. Gwizdz, "The Constitutional Review of Laws in Poland," Polish Report to the lst Congress of the International Association of Constitutional Law held in Belgrade, September 1983 (unpublished); Uppsala Report by P. Nikolić, *supra* note 13, § III.2.C; Uppsala General Report by M. Cappelletti, *supra* note 8, § 2.

[16] *See* the enlightening Uppsala Reports by J. Carpizo & H. Fix Zamudio on Latin America and by B. O. Nwabueze on Africa, *supra* note 8.

[17] The South African story of the struggle of a court against some excesses of an illiberal regime is most significative and deserves to be recounted in some detail. The "constitutional crisis" of that country can be traced back to the 1952 decision of the South African Supreme Court in the case of *Harris and Others v. Minister of the Interior,* 1952(2) S.A. 428 (A.D.), also known as the *Vote* case. In the decision, the Court held that Act 46 of 1951 (the Separate Representation of Voters Act) was unconstitutional. The Act had the basic effect of disenfranchising the Cape Coloureds and the Court ruled that this was violative of certain "entrenched" sections in the South African Constitution (the South Africa Act of 1909), particularly section 35 which provided that "no law . . . shall disqualify any person . . . who is or may become capable of being registered as a voter from being so registered . . . by reason of his race or colour only, unless [passed by a two-thirds majority

of the Senate and of the House of Assembly in joint session]". At the time, the judicial review authority of the Supreme Court with respect to the "entrenched" sections was the subject of great debate and the decision by Chief Justice A. van de Sandt Centlivres (which was declared by at least one commentator – the then Dean of Harvard Law School, Erwin Griswold – to rank among the best decisions in constitutional law, *see* Griswold, "Comment: The 'Coloured Vote Case' in South Africa," 65 *Harv. L. R.* 1361, 1374 (1953)) created quite a controversy in South Africa. In the decision, Chief Justice Centlivres declared that the Court was competent to inquire whether an Act of Parliament had been validly passed: "to hold otherwise would mean that courts of law would be powerless to protect the rights of individuals which were specially protected in the Constitution of this country." *Vote* case at 479. The South African government was not pleased with the *Vote* decision and later in 1952 each House of Parliament by simple majority enacted the High Court of Parliament Act, 35 of 1952, which created a "High Court" of which every member of Parliament was to be a member. The "High Court" was declared to be a court of law and was given the power to review any decision by the Supreme Court which declared an Act of Parliament to be invalid. The "High Court" then proceeded to review the *Vote* case and overruled it holding that the "entrenched" sections of the South African Constitution were no longer binding. The High Court Act, however, was quickly challenged and held by the Supreme Court to be invalid, the judges agreeing that the High Court was simply "Parliament in disguise." *Minister of the Interior v. Harris,* 1952(4) S.A. 769 (A.D.). Finally, in 1955 in another attempt to overcome the judiciary's "unfriendliness," Parliament enlarged the Senate (and the judiciary) and loaded the enlarged Senate with National Party supporters so that at a joint sitting of both Houses of Parliament the Government would have a two-thirds majority of the total membership of both Houses. Senate Act, 53 of 1955. In the following year, the South African Amendment Act, 9 of 1956, was passed at such a joint sitting. This constitutional Amendment considerably revised the entrenched clauses of the Constitution and particularly deleted section 35 relating to franchise rights. The constitutional Amendment also considered judicial review. Section 2 stated in general terms: "No court of law shall be competent to enquire into or to pronounce upon the validity of any law passed by Parliament [except those that effect the "entrenched" sections]." Of course, as previously noted, the entrenched sections had been weakened considerably by the Amendment. The Act enlarging the Senate and the constitutional Amendment were challenged but upheld by the newly expanded judiciary. The Government was thus able to proceed and successfully disenfranchise the Cape Coloureds as it had originally attempted to do in 1951. The uproar caused by these developments was finally put to rest in 1961 when South Africa formally became a Republic. The South African Constitution of 1961 fully reflected the weakened "entrenched" sections and the wording of the 1956 Amendment. This absence of effective judicial review has been retained in South Africa's 1983 Constitution. The new constitution, however, has incorporated some "bill of rights" terminology in its preamble and it might imply a more active constitutional role for the courts. Whether judicial review can become viable again in South Africa remains to be seen.
For further discussion of the South African "constitutional crisis" of the 1950's, *see* H.J. May, *The South African Constitution,* 3rd ed., Capetown, Juta & Co., 1955; B. Beinhart, "The South African Appeal Court and Judicial Review," 21 *Modern L. R.* 587 (1958). For further discussion of judicial review and the constitutional background in South Africa *see* H. Hahlo & E. Kahn, *The South African Legal System and Its Background,* Cape Town, Juta & Co., 1968, 53-63; L.A. Rose Innes, *Judicial Review of Administrative Tribunals in South Africa,* Capetown, Juta & Co., 1963, 1-20. For discussion of South Africa's 1983 Constitution (and further bibliography) *see* L.J. Boulle, *Constitutional Reform and Apartheid,* New York, St. Martin's Press, 1984.

[18] *See* Uppsala Report by Y. Taniguchi on Japan, *supra* note 8.

[19] Perhaps the historical reason for this basic difference, which is reflected in the profound difference between the French and the American versions of "separation of powers" (*see* text accompanying and following note 33, *infra*), lies in the fact that for the American independentists Parliamentary supremacy had the hateful meaning of supremacy of the British Parliament. This might explain why, as Professor Henkin says, the framers of the American Constitution "were not content with democracy, even with representative government, for parliament, too, they had learned, could be despotic." Louis Henkin, *The Rights of Man Today*, Boulder, Colorado, Westview Press, 1978, at 10, quoting *The Federalist* no. 47 (James Madison): "The accumulation of all powers, legislative, executive, and judiciary, in the same hands, whether of one, a few, or many, and whether hereditary, self-appointed, or elective, may justly be pronounced the very definition of tyranny."

[20] The gradual development and the political consequences of the *Parlements*' "power to veto the king's legislation" are described in John P. Dawson, *The Oracles of the Law*, Ann Arbor, The University of Michigan Law School, 1968, 362-371. The veto power was exercised both by a refusal to register the royal enactment on the records of the *Parlement* and by a public protest (*remonstrance*) against a royal act of which the *Parlement* disapproved.

[21] *See, e.g.*, J. P. Dawson, *supra* note 20, at 350-362.

[22] The so-called "principle of the venality of justice" was abolished in France by the Revolution with the law of August 16-24, 1790, tit. II, art. 1. *See* M. Cappelletti & J. Gordley, "Legal Aid: Modern Themes and Variations," 24 *Stanford Law Rev.* 347, 355 (1972).

[23] "By 1750 the Parlements had emerged as an articulate and determined opposition, resisting every effort at moderate reform that successive ministers sought to propose. The Parlements rested their claims on the highest grounds. They invoked the fundamental laws of the kingdom and claimed to be guarding its liberties. . . . [T]heir consistent line of policy was defense of the privileges of the nobility and resistance to all change in administration." J.P. Dawson, *supra* note 20, at 369.

[24] The Revolution promptly dissolved all the *Parlements* and, "[w]hen one considers the savagery of popular feelings against them, it is surprising that so few of their members were actually guillotined, though this is largely because so many escaped into exile." J.P. Dawson, *supra* note 20, at 370. "The highest rate of mortality was suffered by the *parlementaires* in Toulouse (55 executed, i.e., 50 per cent). Bordeaux was next (26 executed, i.e., 20 per cent), and Paris third (28 executed or 14 per cent)." *Ibidem*, at 370, n.22.

[25] *See* Robert Derathé, in Montesquieu, *De l'Esprit des lois*, Tome I, Paris, Editions Garnier Frères, 1973, at LXVII-LXVIII.

[26] *De l'esprit des lois*, book XI, ch. VI (translations in this article, unless otherwise indicated, are by M. Cappelletti).

[27] *Ibidem.*

[28] *Ibidem.* Montesquieu's theory is not without serious ambiguities, however. Contrary to J.-J. Rousseau, who was in favor of a "republican" government – *i.e.*, a government "guided by the general will, which is the law" (Jean-Jacques Rousseau, *Du contrat social ou Principes du Droit politique*, book II, ch. VI, note 1 *et passim*) or, in Montesquieu's definition, a government in which "the people . . . have the sovereign power" (Montesquieu, *op. cit.*, book II, ch. 1) – Montesquieu advocated a moderate, nonabsolute or, as we would call it, constitutional monarchy (book II, ch. IV and book V, ch. XI), while condemning the despotic form of government (book II, ch. V and book V, ch. XIV). Montesquieu's monarchy was characterized by him as the system in which "only one person governs, but

on the basis of fixed and established laws" (book II, ch. 1) and was contrasted to the despotic regime in which "one person decides everything, with no laws and no rule, based merely on his will and caprices" (*ibidem*). Repeatedly, he emphasized that the nondespotic monarch shall be bound "by fundamental laws" (book II, ch. IV; book V, ch. XI) and shall not concentrate in himself the judicial function (book VI, ch. V). In so doing, however, Montesquieu seems to support exactly what the *Parlements* had for a long time been trying to do – to impose, even against the monarch, the superiority of certain, essentially unwritten, "leges generales" or fundamental laws, "fixes et immuables," which, however, due to their essentially fluid and vague nature, were interpreted arbitrarily by those courts of justice of the ancien régime. This doctrine of the superiority of the "fundamental laws of the kingdom" led the *Parlements* to exercise what amounted essentially to judicial review of legislation. *See Judicial Review, supra* note 4, 32-36; and *see* text and note 20, *supra*. There are at least two passages in *L'esprit des lois* where Montesquieu seems to support such judicial function of the *Parlements*: in book III, ch. X, where he complains that in the despotic regime, the "prince" requires absolute obedience to his will and no "remonstrances" are allowed; and in book V, ch. X, where he praises the (moderate) monarchy for the fact that, although the affairs of state are conducted by a person alone, hence more efficiently and promptly than in a "republican" government, such efficiency does not degenerate into rashness because state action is bound to respect the laws. Did, then, Montesquieu "repudiate" – or, in fact, contradict – himself? Where is the "real" Montesquieu? How can such passages be reconciled with the continuous spell – that judicial decisions shall not be "arbitrary" (book XI, ch. VI); that in the good monarchy, the judge's virtue in "la médiocrité" (book XX, ch. XIII); that judges are bound to apply rigorously the *loi* ("where it is precise, the judge simply follows it; where it is not, he searches its spirit") (book VI, ch. III); that the judgment shall never reflect the personal opinions of the judge (book XI, ch. VI); that no liberty exists when the judge is also legislator (book XI, ch. VI); in sum, that the judge has to be only the "inanimate" mouth of the law (*ibidem*)? One explanation, of course, could be the influence of the natural law theories dominating all over Europe in the XVII-XVIII centuries, to which Montesquieu paid more than lip service (*see, e.g.*, book I, ch. II; for a learned interpretation *see* R. Shackleton, "Montesquieu in 1948," 3 *French Studies* 299, 309-323 (1949)). Such theories affirmed the existence of certain fundamental, unwritten laws, rooted in human nature or reason, immutable and universal and superior to the positive laws of the time and place. These theories, too, were hardly compatible with a purely mechanical role of the judge, however, lest the obedience to and the enforcement of natural law were to be left to the exclusive concern of the sovereign, as in Hobbes' conception which was harshly condemned by Montesquieu (*see* R. Shackleton, "Montesquieu in 1948," 3 *French Studies* 299, at 310-311 (1949)). Another explanation might be that Montesquieu did not after all attribute too great an importance to the "fundamental laws" in the determination of the role of the judges. One should note that, although even the kings of France admitted the existence of such laws, by which their power was bound, "they had limited the number (of these laws) to two only: the law regulating the succession to the crown (the Salic Law) and the law establishing the inalienability of the royal domain" (thus R. Derathé, *supra* note 25, at 430). This was much too little, of course, to represent the foundation for a system of judicial review of the monarch's legislation. A third explanation, which seems more plausible to me, is that Montesquieu's vision, and defense, of the "moderate monarchy" was one in which the monarch's powers were limited by the "pouvoirs intermédiaires (*see, e.g., De l'esprit des lois*, book II, ch. IV) and most particularly by the nobility (*ibidem*) rather than by the role of the courts. The contrary opinion of Derathé, *cit.*, at XXXI-XXXII, does not seem convincing; even the so-called "éloge de l'état de la robe" by Montesquieu in book XX, ch. XXII is far from signifying what Derathé seems to suggest.

For, as already mentioned, even there Montesquieu magnifies the "médiocrité" and the "suffisance" of the judge "parlementaire," although he pays lip service to the "gloire" of the "corps" as such, a glory, at any rate, which is immediately declared quite inferior to that of the nobility (*ibidem*). And indeed, how prophetic he was in preaching such mediocrity! Implementing the spirit of his recommendation, not many decades later France and, in the wake of France, much of Continental Europe were to introduce a career judiciary, made up of bureaucratic civil servants – the real glory of *médiocrité*. As I have tried to demonstrate elsewhere (*see Judicial Review, supra* note 4, at 60-66, and "The Doctrine of Stare Decisis and the Civil Law," in *Festschrift für Konrad Zweigert*, H. Bernstein, U. Drobnig & H. Kötz eds., Tübingen, Mohr, 1981, 381, 387-393), this social and professional "mediocrity" of the ordinary Continental ("Civilian") judges was to become one of the reasons why they were unsuitable for the challenging role of judicial review of both administrative and legislative action. Hence this was one important reason why special administrative courts had to be created in the 19th century, and special constitutional courts in our century, to fulfill that role. Today, administrative and, even more so, constitutional judges in Europe resemble federal American judges much more than they resemble ordinary Civilian judges.

[29] *De l'esprit des lois*, book XI, ch. VI.

[30] *Ibidem*.

[31] *See* Jean-Jacques Rousseau, *Du contrat social ou Principes du Droit politique*, Book II, Chs. VI-VII *et passim*.

[32] The influence of "the dictates of Montesquieu's logic in producing this strict separation of governmental powers, which was to remain a basic feature of French judicial organization," is affirmed by J.P. Dawson, *supra* note 20, at 376.

[33] The frequently made affirmation that "the Constitution of the United States . . . embodies [Baron de Montesquieu's] idea" of the separation of powers – as one can read, *e.g.*, in P.P. Wiener, ed., *Dictionary of the History of Ideas*, II, New York, Charles Scribner's Sons, 1973, 251 – is, to say the least, of dubious justification. The fact is that "séparation des pouvoirs," French style, is a profoundly different thing from the American version of it, better described as "checks and balances." *See, e.g.*, J.H. Merryman, *The Civil Law Tradition*, 2nd ed., Stanford, Cal., Stanford University Press, 1985, 15-16; Sir Otto Kahn-Freund, "Common Law and Civil Law – Imaginary and Real Obstacles to Assimilation," in M. Cappelletti, ed., *New Perspectives for a Common Law of Europe*, Leyden & Bruxelles, Sijthoff & Bruylant, 1978, 137, at 159. The French (and Continental European) history of *justice administrative* in the 19th century and of *justice constitutionnelle* in our time, on which *see infra* in this Section, would be totally incomprehensible if the above affirmation were correct.

[34] Even legal positivists agree. *See, e.g.*, Alf Ross, *On Law and Justice* Berkeley, University of California Press, 1959, 284 ("no concrete situation allows for only one application of the law"); H.L.A. Hart, *supra* note 2, at 629 ("the existing law imposes only limits on our choice and not the choice itself").

[35] The full text of the *loi* can be found in J.B. Duvergier, *Collection complète des lois etc.*, I, Paris, Guyot et Scribe, 1834, 310-333.

[36] For the French precedents *see Judicial Review, supra* note 4, at 33 and the references in note 30 *ibidem*. Most constitutions of Eastern European and other Socialist countries entrust the role of controlling the constitutionality of legislation to the "Supreme Soviet" or "Popular Assembly" and/or their "praesidiums." The Yugoslav constitutionalist Pavle Nikolić, in his Uppsala Report on the Socialist countries, *supra* note 13, informs us that this "autocontrol, i.e., the control of the constitutionality of legislation entrusted to the legislative body itself," has proved to be "ineffective." This very ineffectiveness has been the main reason which brought Yugoslavia to adopt, and Czechoslovakia and Poland to try

to adopt, a system of judicial rather than political review. *See* Nikolić's Report, § II.2.1, III.2.A-C; and *see* text and notes 13-15, *supra.*

[37] For a brief survey of the development of *justice administrative* in France *see* L.N. Brown & J.F. Garner, *French Administrative Law*, 3rd ed., London, Butterworth, 1983, 28-31.

[38] The influence of the French system of administrative justice outside France is discussed in *id.*, 162-171.

[39] See the Uppsala Report by L. Favoreu, *supra* note 8, § 48; *see also* Jean Rivero, *Le Conseil constitutionnel et les libertés*, Paris, Economica, 1984, 166.

[40] *See* M. Cappelletti, "Nécessité et légitimité de la justice constitutionnelle," in *Cours constitutionnelles, supra* note 7, 461, 464-471 (hereinafter cited as "Nécessité et légitimité").

[41] *See* note 5, *supra.*

[42] The developments in France are described in M. Cappelletti, "The 'Mighty Problem' of Judicial Review and the Contribution of Comparative Analysis," 53 *Southern California Law Rev.* 409, 412-421 (1980) (hereinafter cited as "Mighty Problem"); *see also* the Uppsala Report by Favoreu, *supra* note 8, § 15-20.

[43] Uppsala Report by L. Favoreu, *supra* note 8, § 38 *et passim.*

[44] On the "infirmities" of the French system *see* "Nécessité et légitimité," *supra* note 40, 499-501.

[45] Some of the *Conseil Constitutionnel*'s most remarkable decisions are translated in M. Cappelletti & W. Cohen, *Comparative Constitutional Law*, Indianapolis, Bobbs-Merrill, 1979, ch. 3.C and ch. 5.C. (hereinafter cited as *Comparative Const. Law*). *See* L. Favoreu & L. Philip, *Les grandes décisions du Conseil constitutionnel*, 3rd ed., Paris, 1984.

[46] *See Judicial Review, supra* note 4, at 36; J.H. Merryman, *supra* note 33, at 16.

[47] The phrase quoted in the text has a literature of its own, discussing inter alia who deserves paternity of it (De Lolme? Bagehot?). *See, e.g.*, W. Holdsworth, 12 *A History of English Law*, London, Methuen, 1938, reprinted 1966, at 344 n.5; H.J. Abraham, *The Judicial Process*, Oxford University Press, 2nd ed. 1968, at 295.

[48] The *Bonham's* decision, as is well-known, affirmed the judicial power to control the validity of legislation: "for when an act of parliament is against common right and reason, or repugnant, or impossible to be performed, the common law will controul it and adjudge such act to be void." 8 Coke's Reports 118a; 77 Eng. Rep. 652. For a very learned commentary *see* T.F.T. Plucknett, "Bonham's Case and Judicial Review," 40 *Harvard Law Rev.* 30-70 (1926). *See* also *Judicial Review, supra* note 4, 36-41.

[49] In a passage much criticized by John Austin, Blackstone affirmed that natural law, being "superior in obligation" to positive law, "is binding over all the globe, in all countries, and at all times: no human laws are of any validity, if contrary to this; and such of them as are valid derive all their force, and all their authority, mediately or immediately, from this original," so that "we are bound to transgress that human law" which is violative of natural law. *See* William Blackstone, *Commentaries on the Laws of England*, Introduction, Section the Second "Of the Nature of Laws in General" (Oxford, Clarendon Press, 1765, at 41, 43); John Austin, *The Province of Jurisprudence Determined*, 1832, Lecture V (London, Weidenfeld & Nicolson, 1954, at 184-186). Nevertheless, Blackstone also affirmed that the Parliament's power is "so transcendent and absolute, that it cannot be confined . . . within any bounds. . . It hath sovereign and uncontrollable authority in making, confirming, enlarging, restraining, abrogating, repealing, reviving, and expounding of laws, concerning matters of all possible denominations, ecclesiastical, or temporal, civil, military, maritime, or criminal: this being the place where that absolute despotic power, which must in all governments reside somewhere, is entrusted by the constitution of these kingdoms. . . It can, in short, do everything that is not naturally impossible; and therefore some have not scrupled to call its power, by a figure rather too bold, the omnipotence of

parliament. True it is, that what the parliament doth, no authority upon earth can undo. . . So long, therefore, as the English constitution lasts, we may venture to affirm, that the power of parliament is absolute and without control." Blackstone, *Commentaries*, Book I, Ch. II, at 160-162; *see* also Introduction, Section the Third "Of the Laws of England," at 90-91, where Blackstone rejects the possibility for the judges to set aside parliamentary laws even if these laws command "a thing to be done which is unreasonable." "I know of no power," he says, "that can control" such laws; and to maintain that the judges "are at liberty to reject" them, "[would] set the judicial power above that of the legislature, which would be subversive of all government." As Pound rightly noted, when Blackstone "comes to apply [his theory of natural law] to legislation, he retracts." R. Pound, "Common Law and Legislation," 21 *Harv. L. Rev.* 383, at 392 (1908).

[50] John Locke, *The Second Treatise of Government*, Ch. XII, § 144, 149 *et passim.*

[51] *The Second Treatise*, Ch. XII, § 145-148 *et passim.*

[52] *The Second Treatise*, 52 Ch. XI, § 134, Ch. XIII, § 149, *et passim.*

[53] *The Second Treatise*, Ch. XIX, § 212. It might be true, however, that because the legislative power, and more generally "the power to govern," was seen by Locke as "the pronouncing and enforcing of a law, the law of nature which is the law of reason," he saw a "judicial" feature inherent in and pervading that power, as noted by Peter Laslett, "Introduction" to John Locke, *Two Treatises of Government*, Cambridge, At the University Press, 1960, 96, 107; *see, e.g., The Second Treatise*, Ch. VII, § 88-89, Ch. XI, § 136.

[54] *See, e.g., The Second Treatise*, Ch. IX, § 124.

[55] *See, e.g., The Second Treatise*, Ch. XI, § 135.

[56] *See* note 49, *supra.*

[57] *See, e.g.,* G. Wintertorn, "The British Grundnorm: Parliamentary Supremacy Re-Examined," *Law Quarterly Review*, 1976, 591-617.

[58] *See, e.g.,* C. Sasse & H.C. Yourow, "The Growth of Legislative Power of the European Communities," in T. Sandalow & E. Stein, eds., *Courts and Free Markets*, I, Oxford, Clarendon Press, 1982, 92-126. *See* generally E. Stein, P. Hay & M. Waelbroeck, *European Community Law and Institutions in Perspective*, Indianapolis, Bobbs-Merrill, 1976; Commission of the European Communities, *Thirty Yecrs of Community Law*, Luxembourg, Office for Official Publications of the European Communities, 1983.

[59] *See Comparative Const. Law, supra* note 45, at 137; the quotation is from Bulmer v. Bollinger, (1974) 2 All E. R. 1226.

[60] The principle was affirmed for the first time in the historic 1963 decision by the European Court of Justice (ECJ), van Gend en Loos v. Nederlandse Administratie Belastingen, Case 26/62, (1963) ECR 1. *See, e.g.,* L.N. Brown & F.G. Jacobs, *The Court of Justice of the European Communities*, 2nd ed., London, Sweet & Maxwell, 1983, at 162.

[61] *See Comparative Const. Law, supra* note 45, at 137. *See also* M.P. Furmston, R. Kerridge & B.E. Sufrin, eds., *The Effect on English Domestic Law of Membership of the European Communities and of Ratification of the European Convention on Human Rights*, The Hague, Nijhoff, 1983, 1-246.

[62] The first affirmation of the preeminence doctrine can be found in another historic decision of the ECJ, Costa v. ENEL, Case 6/64, (1964) ECR 585. On the gradual acceptance by most of the national courts of the doctrine of the supremacy of Community law *see* "Mighty Problem," *supra* note 42, at 424-426; R. Kovar, "The Relationship Between Community Law and National Law," in *Thirty Years of Community Law, supra* note 58, at 109.

[63] *See* the ECJ decision in Simmenthal, Case 106/77, (1978) ECR 629.

[64] *See, e.g.,* L.N. Brown & F.G. Jacobs, *supra* note 60, at 281-285; decision of the ECJ Da Costa en Schaake, Cases 28-30/62, (1963) ECR 31. Generally on the role of the ECJ *see* M. Cappelletti & D. Golay, "The Judicial Branch in the Federal and Transnational Union:

Its Impact on Integration," in M. Cappelletti, M. Seccombe & J. Weiler, eds., *Integration Through Law: Europe and the American Federal Experience*, Vol. 1, Book 2, Berlin & New York, de Gruyter, 1986, 261-351.

[65] C.L. Black Jr., "Is There Already a British Bill of Rights?," 89 *Law Quarterly Rev.* 173 (1973).

[66] The acceptance, first limited to a period of three years, was since regularly renewed; the last renewal occurred in 1981 for a period of five years. *See generally* A.Z. Drzemczewski, *European Human Rights Convention in Domestic Law*, Oxford, Clarendon Press, 1983, 177-187, 362-363.

[67] For brief surveys *see, e.g.*, Rosalyn Higgins, "The European Convention on Human Rights," in T. Meron, ed., *Human Rights in International Law*, II, Oxford, Clarendon Press, 1984, 495-549; A.H. Robertson, *Human Rights in the World*, 2nd ed., New York, St. Martin's Press, 1982, 80-117. The Convention has been ratified by the following countries (in parenthesis is, first, the year of ratification, and second, the year, if any, of the acceptance, ever since uninterrupted, of the optional clause of Art. 25): Austria (1958; 1958); Belgium (1955; 1955); Cyprus (1962); Denmark (1953; 1953); France (1974; 1981); Federal Republic of Germany (1952; 1955); Greece (1974; 1985); Iceland (1953; 1955); Ireland (1953; 1953); Italy (1955; 1973); Liechtenstein (1982; 1982); Luxembourg (1953; 1958); Malta (1967), Netherlands (1954; 1960); Norway (1952; 1955); Portugal (1978; 1978); Spain (1979; 1981); Sweden (1952; 1952); Switzerland (1974; 1974); Turkey (1954); and the United Kingdom (1951; 1966). Of these countries, only three – Cyprus, Malta and Turkey – have not yet accepted the optional clause of Art. 25. For further information concerning also the ratifications of the Convention's Protocols *see* A.Z. Drzemczewski, *supra* note 66, at 362.

[68] On the impact of the European Convention on the United Kingdom *see* A.Z. Drzemczewski, *supra* note 66, 177-187, with reference to a number of cases; M.P. Furmston, R. Kerridge & B.E. Sufrin, *supra* note 61, 247-428.

[69] On the authority of and compliance with the decisions of the adjudicatory organs of the European Convention on Human Rights in the various member states, *see* A.Z. Drzemczewski, *supra* note 66, at 260-325. With particular regard to the United Kingdom, *see* the recent comments by the Chairman of the Law Commission for England and Wales, the Honourable Mr. Justice Gibson, "Legal Procedure: Access to Justice: 1883 to 1983," in 9 *Dalhousie Law Journal* 3, 27-28 (1984):

> The United Kingdom has been adjudged to be in breach of its obligations under the European Convention in a number of cases which I can only call large. Government has, no doubt, found these events both surprising and embarrassing. Breach was established in a case about the working of the closed shop in our nationalized railways for which damages . . . and costs . . . were awarded to three claimants. There have been cases . . . about immigration. Breaches have also been established on individual petitions in other contexts, such as the use of corporal punishment in a school without the consent of the parents, the censorship of mail by prison authorities and the refusal of permission for a prisoner to seek legal advice, and the working of the common law of contempt against *The Sunday Times* newspaper in its investigation and reporting of the Thalidomide case. In one case, *Eire v. U.K.* the proceedings were between two parties of the Convention. It was alleged that the authorities in Northern Ireland had inflicted torture on Republican prisoners by using a number of interrogative devices, such as wallstanding, subjection to noise, and deprivation of sleep. The (European) court held that the techniques did not amount to torture, but were inhumane and degrading treatment in breach of Article III.
>
> The response of the government to these and other decisions has been to discontinue the offending practice, often before decision by the court, and, when necessary, to

change the relevant law, such as prison rules. There has been no sign of an intention to defy the decision of the court, although not everyone agrees with the interpretation of the Convention by the various majorities in the Court of Strasbourg. The United Kingdom could denounce the Convention on six months' notice under Article LXV. . . It has not done so and such a step is exceedingly improbable.

[70] *See* "Mighty Problem," *supra* note 42, at 424-430.

[71] *See* the Uppsala Report by Eivind Smith on the Scandinavian countries, *supra* note 8; *see also, e.g.*, G. Hahn, "Verstärkter Grundrechtsschutz und andere Neuerungen im Schwedischen Verfassungsrecht," *Archiv des öffentlichen Rechts*, 1980, 400-422.

[72] *See* the Uppsala Report by John D. Whyte on the Common Law countries, *supra* note 8; *see also, e.g.*, A. Bayefsky, "Parliamentary Sovereignty and Human Rights in Canada: The Promise of the Canadian Charter of Rights and Freedoms," *Political Studies*, 1983, 239; J.B. D'Onorio, "Le repatriement de la Constitution canadienne," *Revue internationale de droit comparé*, 1983, 69 (esp. at 100-101 on the serious consequences of the "notwithstanding clause" of Section 33 of the Charter).

[73] An encouraging sign is the recent landmark decision by the Inter-American Court which unanimously held that a Costa Rican law requiring the compulsory licensing of journalists violated the freedom of opinion and expression guaranteed by Art. 13 of the American Convention on Human Rights, *Schmidt v. Costa Rica* (Inter-American Court of Human Rights, Judgment of Nov. 14, 1985). *See generally*, T. Buergenthal, R. Norris & D. Shelton, *Protecting Human Rights in the Americas*, Strasbourg, Engel Publ., 1982; T. Buergenthal, "The Inter-American System for the Protection of Human Rights," in T. Meron, *supra* note 67, II, 439-493; *id.*, "The American and European Conventions on Human Rights: Similarities and Differences," 30 *American Univ. Law Rev.* 155 (1981); H. Gros Espiell, "The Organization of American States (OAS)," in K. Vasak, ed., *The International Dimensions of Human Rights*, II, Westport, Connecticut, Greenwood Press, 1982, 543-565; P. Sieghart, *The International Law of Human Rights*, Oxford, Clarendon Press, 1983, 401-414. *See* also Inter-American Commission on Human Rights, *Ten Years of Activities 1971-1981*, General Secretariat Organization of American States, Washington, D.C., 1982 (at 11-13 discussing the newly created Inter-American Court of Human Rights); Organization of American States, Inter-American Court on Human Rights, *Annual Report of the Inter-American Court of Human Rights 1984*, Washington, D.C., General Secretariat OAS, 1984 (a sadly impressive document of the, as yet, notwithstanding the encouraging 1985 *Schmidt* case previously discussed in this note, extremely limited activity of the Inter-American Court).

[74] For a more elaborate discussion I shall refer to my studies "The Law-Making Power of the Judge and Its Limits: A Comparative Analysis," 8 *Monash University Law Rev.*, 15, esp. 51-58 (1981) (hereinafter cited as "Law-Making Power"); "Nécessité et légitimité," *supra* note 40, 475-483.

[75] *See* generally H. Fix Zamudio, *Veinticinco años de evolución de la Justicia Constitucional 1940-1965*, Mexico, UNAM, 1968, esp. ch. 2.

[76] *See* the Uppsala Reports by Nwabueze, *supra* note 16, at 18-23, and Carpizo & Fix Zamudio, *supra* note 16, § 61-69, 91, 95, 110 *et passim*. Professor Henkin points out "what today might seem a striking but happy omission" of the U.S. Constitution which "does not provide for its suspension, or for government by decree even in emergency." "Only the privilege of the writ of habeas corpus may be suspended." L. Henkin, *supra* note 19, at 13, 150 n.31.

[77] Guido Calabresi, *A Common Law for the Age of Statutes*, Cambridge, Mass., Harvard University Press, 1982.

[78] *See e.g.*, Robert Dahl, "Decision-Making in a Democracy. The Supreme Court as a National Policy-Maker," 6 *Journal of Public Law*, 279, esp. at 284-285 (1957).

[79] *See* "Law-Making Power," *supra* note 74, espec. at 42-46, 54-57. Even though their profession and role might to some extent insulate the judges from society, their very activity "forces the judges down to realities, since they are called to decide cases involving live persons, concrete facts and actual problems of life." *Id.*, at 57.

[80] Criticism, of course, is facilitated by the fact that in our societies the most important judgments and their reasons are published; it is further facilitated in those countries where dissenting and concurring opinions are also published. *See* the comparative study by K.H. Nadelmann, "The Judicial Dissent: Publication v. Secrecy," 8 *Am. J. Comp. Law* 415 (1959).

[81] *See* M. Cappelletti, "Who Watches the Watchmen? A Comparative Study on Judicial Responsibility," 31 *Am. Journal of Comparative Law* 1 (1983).

[82] Applying the teaching of the great historian-philosopher Vico, "verum ipsum factum," the comparativist "speculates" on the significance of facts, trends, and developments, not of abstractions. Giambattista Vico, *Principj di Scienza Nuova* (1744); English transl. by T.G. Bergin and M.H. Fisch, *The New Science of Giambattista Vico*, Ithaca, N.Y., Cornell University Press, 1948. Comparative analysis, of course, is not only comparison of contemporary laws and legal systems but also analysis of their evolution and trends. History, in other words, is an essential component of comparative analysis.

[83] *See, e.g.*, Alexander M. Bickel, *The Least Dangerous Branch*, Indianapolis, Bobbs-Merrill, 1962, 25-27 *et passim*.
It is frequently said that modern constitutional adjudication, while potentially a powerful instrument for the protection of traditional political rights and values, has no potential for also being or becoming so for the protection and enforcement of the "new" social and economic rights. For these rights, unlike the traditional ones, usually require affirmative state action, and the judicial mandate is said to be powerless to determine such action. This is only a half truth, however. Greater difficulties are certainly encountered and greater restraint is advisable when courts, ascertaining the illegitimacy of governmental inaction, command the government to do something, with all the economic and other implications thereby involved, than when they simply declare the illegitimacy of an existing governmental act. Comparative study demonstrates, however, that there are many ways for the courts to intervene even in this more difficult area. A most recent example is provided by the Burger Court, certainly not an activist court in the sphere of socio-economic rights. *See* Ake v. Oklahoma, 105 S. Ct. 1087 (1985), in which, with only one dissent, the Court held that the states must provide indigent criminal defendants with free psychiatric assistance in preparing an insanity defense if the defendant's sanity at the time of the crime is likely to be a significant factor at trial.

[84] *See, e.g.*, Martin Shapiro, *Freedom of Speech: The Supreme Court and Judicial Review*, Englewood Cliffs, N.J., Prentice Hall, 1966, 37 *et passim*; J.H. Ely, *Democracy and Distrust*, Cambridge, Mass., Harvard University Press, 1980. The central thesis of Dean Ely's book is its advocacy of a "representation reinforcing" approach; this approach emphasizes the role of judicial review as an instrument for the preservation of a fair, open political process and for the correction of "malfunctions" of this process which would impair the effective participation of minorities.

[85] *See* M. Cappelletti & B. Garth, "Access to Justice: The Worldwide Movement to Make Rights Effective," in M. Cappelletti & B. Garth, eds., *Access to Justice*, Vol. I: *A World Survey*, Milan & Alphen aan den Rijn, Giuffrè & Sijthoff and Noordhoff, 1978, 3-124. *See generally* M. Cappelletti, ed., *Access to Justice and the Welfare State*, Alphen aan den Rijn & Bruxelles, Sijthoff & Bruylant, 1981.

[86] *See, e.g.*, L. Henkin, ed., *The International Bill of Rights. The Covenant on Civil and Political Rights*, New York, Columbia University Press, 1981; *id.*, *supra* note 19, at 89-101. Another major development of our time, the drive against colonialism, which brought

about the creation of dozens of new nations in the post-World War II years, can be seen as an expression of the trend mentioned in the text. As Professor Henkin says, it was that drive that "brought a mass of new states and governments that looked to the idea of human rights to achieve 'self-determination' and the elimination of racism." *Id., supra* note 19, at 18.

[87] *Cf. id., supra* note 19, at 101-102, 107-113; L. Henkin, "International Human Rights as 'Rights'," in *Nomos XXIII: Human Rights*, J.R. Pennock & J.W. Chapman, eds., New York, New York University Press, 1981, 257-280; L.B. Sohn, "Human Rights: Their Implementation and Supervision by the United Nations," in T. Meron, *supra* note 67, II, 369-401.

[88] Hume, *An Enquiry Concerning Human Understanding*, Sect. XII, Part I (in *Enquiries Concerning the Human Understanding and Concerning the Principles of Morals* by David Hume, reprinted from the Posthumous Edition of 1777, L. A. Selby-Bigge editor, 2nd ed., Oxford, Clarendon Press, 1902, reprint 1963, at 149). Although teaching that human knowledge is limited to experience of ideas and impressions, and excluding the possibility of any definitive verification of their truth, Hume severely condemned what he called "excessive" or "Pyrrhonian" skepticism:

> For here is the chief and most confounding objection to excessive scepticism, that no durable good can ever result from it. . . (The Pyrrhonian skeptic) must acknowledge . . . that all human life must perish, where his principles universally and steadily to prevail. . . (However,) nature is always too strong for principle. . . (T)he first and most trivial event in life will put to flight all (the Pyrrhonian's) doubts. . . When (the excessive skeptic) awakes from his dream, he will be the first to join in the laugh against himself, and to confess, that all his objections are mere amusement, and can have no other tendency than to show the whimsical condition of mankind, who must act and reason and believe; though they are not able, by their most diligent enquiry, to satisfy themselves concerning the foundation of these operations, or to remove the objections, which may be raised against them. (*Ibidem*, Sect. XII, Part II, at 158-160.)

I wonder how much contemporary legal writing would change its theme and tone and how much intellectual energy and talent would be put to better use if Hume's teaching were learned.

[89] L. Henkin, *supra* note 19, espec. at 43-55, 156-161.

[90] Henkin's documentation focuses on, without being limited to, the United States (for an account of developments in Europe *see Comparative Const. Law, supra* note 45, chs. 6-12). The "rights explosion" since World War II is described as "impressive":

> The Fourteenth Amendment was held to have incorporated, and rendered applicable to the states, the principal provisions of the Bill of Rights – freedom of speech, press, assembly, religion, the security of the home and the person, safeguards for those accused of crime. . .
>
> Even more impressive has been the expansion of our eighteenth-century rights in conception and content. We have opened our Constitution to every man and woman, to the least and the worst of them. We have opened it also to new rights and to expanded conceptions of old rights. . . . We safeguard not only political freedom but also, in principle, social, sexual, and other personal freedom, privacy, autonomy,. . . All racial classifications are suspect and sharply scrutinized. . . There has been a fundamental and, I believe, irreversible transformation in the status of women. . . The poor, too, have rights of equal protection, and the state cannot offer important rights – a criminal appeal, a divorce – for pay without making them available gratis to those who cannot pay. . . Other once-closed categories are open: prisoners now have rights, as do military personnel, mental patients, pupils in the schools, and children independently of their parents.

> The courts also found new rights, for example, a right to travel, abroad as well as interstate. . . [They] have found an area of fundamental individual autonomy [called "privacy"], invasions of which will be . . . invalidated unless they serve a compelling state interest. . . (L. Henkin, *supra* note 19, 43-44, 46.).

[91] The undeniable fact that judicial review in America played a modest, at times even a negative role in the protection of civil liberties until a very few decades ago, is often indicated as evidence of the democratic deficit of the institution itself. *See, e.g.*, Peter Railton, "Judicial Review, Elites, and Liberal Democracy," in *Nomos XXV: Liberal Democracy*, J.R. Pennock & J.W. Chapman eds., New York, New York University Press, 1983, 153-180, and the literature mentioned therein.

[92] Luigi Pirandello, *Six Characters in Search of An Author*, Act. III:

> . . .never do people reason so much and become so introspective as when they suffer; since they are anxious to understand . . .whether it is just or unjust that they should have to bear it. On the other hand, when they are happy, they take their happiness as a matter of course, and do not analyze it, just as if happiness were their natural right.

[93] A typical illustration is provided by the contributions in the volume *Natural Law and Modern Society*, Cleveland & New York, The World Publishing Co., 1962. *See also, e.g.*, L. Henkin, *supra* note 19, at 19. *Cf.* R.A. Dworkin, "Natural Law Revisited," 34 *Univ. of Florida Law Rev.* 165-188 (1982).

[94] *See* my studies "Judicial Review in Comparative Perspective," 58 *California Law Rev.* 1017, at 1017-1020, 1032-1033 (1970); "The Significance of Judicial Review of Legislation in the Contemporary World," in *Ius Privatum Gentium, Festschrift für Max Rheinstein* (E. von Caemmerer, S. Mentschikoff & K. Zweigert eds.), Tübingen, Mohr, 1969, 147, at 155-162. *See* also Henkin, *supra* note 19, at 5-23, 148-152, espec. at 19-23.

[95] *See* Jean Rivero, "Rapport de synthèse," in *Cours constitutionnelles, supra* note 7, 517, at 525-526.

# Legal Analysis of Economic Policy

TERENCE C. DAINTITH
Dundee/Florence

## Introduction

### 1. The systemic approach to public law

A leading recent text on British constitutional law, having listed a number of measures adopted by the Conservative and Labour Governments between 1970 and 1977 for the furtherance of their economic policies, concluded:

"Potentially some of these developments were of considerable constitutional interest; but so many of them were influenced by economic and political vicissitudes that the most appropriate commentator was the journalist."[1]

The feeling here expressed, that the writer on public law matters should deliberately distance himself from the hurly-burly of economic events and political action and conflict, and should draw a picture of constitutional and administrative law which is purified of non-legal, short-run influences, is one which has commanded wide acceptance in both academic and practising legal circles in the United Kingdom.[2] There are good intellectual, pedagogical and professional reasons underlying this attitude. The absence of any formal and comprehensive written constitution for the United Kingdom compels a vigorous effort of systematization and abstraction if the constitutionally significant features of a long, and continuing, political and legal history are to be apprehended at all, and at the same time eliminates any risk that such a systematic treatment of the constitution will amount to no more than an arid textual commentary without reference to real life. A more mundane, but equally important consideration is that the lawyer must defend his *métier*: with few authoritative texts on which to deploy his recognized interpretative skills, he runs the risk, in dwelling too long on "economic and political vicissitudes" in his treatment of the constitution, of being mistaken not for a journalist but — a worse fate, perhaps — for a political scientist. In the result, the

mainstream of British writing by lawyers about the constitution analyses constitutional principles and structures largely to the exclusion of any consideration of the activities in which the organs of government are engaged or the purposes with which they pursue them – an interesting exception being the area of governmental activity which bears most directly upon personal liberty, the actions of the police. Certainly, innumerable examples of the behaviour of constitutional organs are cited, but always with the objective of illustrating a confirmation of or a change in the constitutional system, rather than the legal framework of the particular activity of government from which the example is drawn.

The merits of this approach – the distillation of a reasonably coherent system of public law out of a sprawling mass of political and legal events and behaviour – have offsetting disadvantages. Constitutions must reflect the changing powers and responsibilities of government if they are not to risk irrelevance and displacement, whether by peaceful or less than peaceful means. One of the strengths, it may be argued, of the British Constitution is an inherent flexibility and responsiveness which makes it unlikely ever to suffer the fate of the Constitution of the French Fourth Republic; but the practitioner of the systemic approach to constitutional law, dismissing the current activities of government as mere ephemerides, may be slow to appreciate, and reluctant to acknowledge, the flexible response of our constitution to such facts and events and their consequent potential for engendering constitutional changes of continuing importance.

A visible aspect of this approach in the literature is an attitude of suspicion towards the very existence of governmental power. In general, the emphasis is still placed heavily on analysis of the capacity of the law and legal institutions to restrain governmental action, rather than on the exploration of their ability to further governmental purposes. Indeed, the very possibility that law may be shaped (or reshaped) more by government policy than policy is constrained by existing law seems hard for some to accept: administrative law, for one of our leading modern writers, has for its only object "to keep the powers of government within their legal bounds" and to prevent "powerful engines of authority" from "running amok".[7] The more dangerous invisible aspect is the failure to see the constitutional problems of the present which arise from the constant adaptations of the instruments of governmental power and action in order to discharge the current range of governmental responsibilities amid the economic and political vicissitudes of the hour. I would argue, for example, that government's use of its contractual powers to influence private and public sector behaviour was a significant constitutional problem long before their deployment in the contentious context of pay restraint lifted them to the status of a first-rank political issue and won them a place in the constitutional law literature.[8]

## 2. An alternative "policy" approach

Implicit in these criticisms is the idea that a truer picture of current constitutional and public law problems may be obtained by exploring the material hitherto set to one side by the mainstream approach, that is to say, aims and activities of government. I do not suggest that such an exploration can be a substitute for traditional systematic analysis; its role is rather a complementary and corrective one. It would start not from constitutional principles, but from government policy, and ask what demands government makes of the constitutional and legal systems in seeking to achieve its objectives, how those systems respond to those demands, and what problems those responses create for government and for the citizen. This approach might be expected to produce a profile of constitutional problems differing, at least in some significant respects, from that yielded by the systemic approach of mainstream literature. One simple and already familiar example that could be documented in this way is that of government rule-making: unresolved problems of parliamentary control of delegated legislation dwindle into insignificance beside those revealed by the government's modern preference for informal rule-making.

The problems of such an approach to public law scholarship are considerable. How is the vast and shifting mass of data to be organized? How is the investigator to deal with bewilderingly frequent changes in government policy, and in the administrative apparatus used to carry those policies into effect? How are baselines to be drawn for the observation of trends in government activity? This paper seeks to contribute to the study of government policy by the public lawyer, by furnishing some elements of a system for the description and study of policy which can be related to the structure of the constitution and of the legal system. Such a system should ideally be applicable to any country which accepts the idea of purposeful governmental activity within a legal framework, and within such a country, to the analysis of any field of government policy. In the present paper, however, the system is elaborated with reference only to the United Kingdom.

Moreover, within the sphere of governmental activity in the United Kingdom, the ideas here have been worked out in the course of extended study of the economic policy of the period since 1945. The argument in the paper will be couched in terms of economic policy, will draw on the literature of applied economics, and will use illustrations from the economic sphere. My view is that the same structure should be applicable, perhaps with minor modifications, to other fields of policy, but I do not claim to have demonstrated that this is so. By itself, however, the economic policy area is broad and overlaps substantially with other areas of policy, such as social policy and defence policy: conclusions reached through study in this field may claim some presumptive general validity.

## Government policy: components, context, actors

### 1. Components

The term "government policy" can be used narrowly, so as to refer only to *statements* by government about the objectives it sets itself and the means it proposes to use for attaining them. Alternatively, it may bear a broader meaning, and be used to refer to the whole of the *purposeful activity* of government. Such activity forms part of a complex process of interaction between government and other actors in the national and international social and economic environment. In this process government seeks to modify the behaviour of those actors by reference to determined objectives, which are themselves continually amended in response to changing circumstances and to feedback from affected actors.[5]

This concept of policy as purposeful activity forms my starting point. The need is to analyse the activity in a way which shows as clearly as possible its relationship with legal operations: the making, application, and interpretation of law. Here I attempt a partial analysis which concentrates on the implementation phase of economic policy: the process of formulation and amendment of policy objectives is not systematically discussed. For my present purpose a fruitful line of approach is suggested by the analysis originated by Tinbergen.[6] At its core is a clear and rigid distinction between objectives of policy and instruments of policy. Criticism by political scientists[7] of this ends-means distinction, as not reflecting the blurriness of the policy-making process or the tendency of politicians to attach electoral and other values to instruments as such, without regard for the objectives they serve, does not detract from the utility of the approach as a means for clarifying the role of law in the implementation of economic policy. Tinbergen's aim, in fact, was to provide tools for the mathematical modelling of economic policy and the prediction of outcomes, a quest pursued with increasing sophistication by some of his followers.[8] Others, like Kirschen,[9] have elaborated Tinbergen's approach as a framework for orderly comparison of the policies actually pursued by governments of both capitalist and socialist countries. In formulating a legally-relevant analysis of economic policy I have drawn heavily on Kirschen, so that it may be helpful, in order to avoid confusion, to set out first the elements of his system, so that the nature of my own adaptations can be clearly seen.[10]

Kirschen distinguishes, as elements of policy, aims, objectives, instruments, and measures. *Aims* are a series of abstract concepts which represent the desirable ends of government action: economic welfare, physical security, equality, liberty, and so on. *Objectives* are "the economic translations of political aims into concepts which can be given some quantification".[11] In his two large collaborative works Kirschen offers rather different lists of objectives, but the nature of his concept can be adequately

conveyed by listing a selection: full employment; price stability; expansion of production; improvement in the distribution of income and wealth; reduction of regional disparities; protection of specific branches of industry. *Instruments* are the means by which objectives are pursued, and are grouped into such categories as money and credit; budgetary and fiscal; direct controls, such as rationing or price controls; and institutional changes, such as the nationalization of an industry. Finally, a *measure* is the use of a particular instrument at a particular time in order to promote one or more objectives.

In adopting this taxonomy I use the terms *aims*, *objectives*, and *instruments* in the same sense as Kirschen. *Aims* are briefly discussed in the next paragraph. The concept of *objectives* allows considerable freedom as to the degree of precision with which objectives are specified. Most writers on economic policy confine their attention to the classical quartet of general macro-economic policy objectives: full employment, price stability, external balance, economic growth.[12] The lawyer may, however, find it helpful to break down these general ideas into a series of more specific objectives to which particular instruments and measures can be more certainly related. If he wants to analyse policy "areas" such as energy policy, manpower policy, and so on, he will need to analyse their content in terms of precise objectives in this way.[13] The concept of *instruments* is explored in detail later in this paper, but it should be made clear now that the concept is an economic, not a legal one. Legal rules and other acts are thus not themselves instruments, but may be *measures*, which I prefer to define as the *acts through which instruments are brought into operation*. Such acts are usually "legal", in that they are recognizable as formal elements of the legal order; but they need not necessarily be so. An agricultural subsidy is an instrument: the Act authorizing its payment, the statutory instrument under which it is administered, the making of payment to the farmer (whether analysed as contract or as gift), are all (legal) measures. A restriction of bank credit is likewise an instrument: the Bank of England "request" bringing it into effect is a measure, but has no formal legal status.

## 2. Context

The concept of *aims* is nebulous, but valuable in that it reminds us that economic welfare is not the only end for which government was created, and prompts the question of how we relate to the categories of economic policy the means of pursuit of other aims such as equality and liberty. There are certain economic *objectives* whose attainment may promote such aims as well as that of economic welfare, but it is clear that such "goods" as full employment and price stability are neither sufficient nor, some would

add, necessary to a just and free society. At the level of objectives, however, it might well be possible to specify a series of concepts through which such aims are concretized in a given society, and which would represent the political or constitutional counterpart of the economic objectives detailed by Kirschen. In the United Kingdom, sample objectives of this type might be formulated as "democratic endorsement of public spending and taxation decisions", or "performance of (certain) governmental functions within a legal framework supervised by an independent judiciary". Similarly, one might pursue the parallel further in seeking to identify *instruments* for the achievement of these objectives, to be found in procedural or substantive constitutional principles which, in the United Kingdom, would probably be expressed in the form of constitutional conventions, common law rules, and canons of judicial interpretation.

For the purposes of this discussion I do not think it necessary to elaborate fully a parallel structure or structures of this type. The important point is to be aware that the pursuit of aims other than economic welfare engenders values, principles, and rules which bear upon governmental choice at all three operative levels of the Kirschen classification: choice of economic objectives, choice of instruments for their achievement, choice of measures for the operationalization of those instruments. Many of these values, principles and rules form part of the constitutional and legal system. It is the interaction between these elements of public law and of economic policy that defines and justifies the public lawyer's interest in economic policy, for we may postulate that not only do the constitutional and legal systems shape economic policy choices, but that the pursuit of economic policy objectives may also demand and secure changes in these systems. The Kirschen taxonomy provides part of the framework for a systematic discussion of these influences.

A related set of ideas that need to be taken into account in any development of the Kirschen taxonomy for legal purposes are those revolving around the concept of a State's "economic order". Such orders may be labelled as "laisser faire", "social market", "collectivist", "corporatist" and so on, terms designed to express fundamental orientations of economic activity. In recent years this concept has been most intensively developed by German scholars of the theory of economic policy, who commonly distinguish between "ordering policies", which alter the shape of the framework within which economic activity is carried on, and "process policies", which operate in the free space within that framework.[14] This space/framework distinction has been traced back to the English classical economists,[15] and is reflected, in the Tinbergen/Kirschen analysis, in their distinctions between "reforms", "qualitative policies" and quantitative policies", and between instruments that do, and do not, make "institutional" changes.[16] In so far as the economic order of State is expressed in its formal constitution – as is the case in the Federal Republic of

Germany – the distinction between space and framework, ordering policies and process policies, institutional and non-institutional instruments, is clearly a vital one for public lawyers. In the United Kingdom, with its flexible constitution which is largely empty of substantive principles, the distinction does not have the same normative weight: institutional change is, in legal terms, as easy to effect as non-institutional change.[17] Its value lies rather in reminding us, once more, that legal rules may operate not only as agents of change ("measures" in the Kirschen terminology), but also as constituents of the economic framework within which change takes place and which may itself be subject to change. The institution of property, in the United Kingdom largely defined and protected by the common law, is one vital element of such a framework. The relevant common law rules cannot be assimilated to any element of Kirschen's taxonomy, but help to define the space within which that taxonomy operates. And as we shall now see, while they may be changed by the operation of the economic policy process, this is not their only source of development.

## 3. Actors

Having situated this analysis in its proper context it is still necessary to provide a further element of definition, by way of reply to the question: *whose* economic policy? Obviously I intend to discuss "the government's" economic policy, and by this I mean, primarily, the activity in this field of the central government, composed of ministers of the crown and their departments, coordinated by the cabinet, guided by the treasury. But even in a unitary and centralized State like the United Kingdom, the term "government" in such a context is not unequivocal.

In the first place, there exists, at any given moment, a diffusion of power among various organs of government. We need to take into account both the territorial diffusion of power and functions among local authorities, and the functional diffusion of power, both to specialist executive organs like the public corporations which are responsible for the nationalized industries and for certain regulatory and public service functions, and to organs outside the executive altogether, such as parliament and the courts. Within the United Kingdom constitutional context, it is of course true that none of these power-holders is in any sense autonomous: Parliament, in the exercise of its sovereign powers, can at any moment withdraw or change their functions or the conditions of their exercise, and the massive (though not unlimited) influence exercisable by the central executive government over the legislative process in Parliament gives it a large measure of ultimate control over the activities of all organs of government in the United Kingdom. This centralization of ultimate power is not,

however, a good reason for treating the economic policy activities of all organs of government in an undifferentiated way. Even within the framework of the executive branch of government, local authorities and public corporations do not stand in a relationship of hierarchical subjection to the departments of central government. They are separate legal entitites, whose status and competences are guaranteed by law (and in the case of local authorities, reinforced by their democratic and representative character), and which are therefore – at any given moment – subject to the control of central government only to the extent provided for by the law or secured by other means (such as financial dependence) against the background of the law. Given however that that control is frequently considerable, and that the economic importance of these institutions is, by some measures, as great as that of central government itself, we shall find that the manipulation of the behaviour of the whole, or particular parts, of the non-central government part of the public sector is a frequently employed instrument of policy. To define "government" in a broad sense, so as to include the whole of the public sector, would be to treat as internal, and therefore beyond our concern, these processes of manipulation, and to neglect the importance of the varying degrees of legal independence possessed by these bodies, which may often make it more appropriate to treat them as subjects, rather than executants, of policy.

Different arguments apply in relation to non-executive bodies. Parliament, in the British system, should not be viewed as the maker or possessor of economic policies. Its role is rather one of scrutiny, discussion and legitimation of policies formed elsewhere. In almost every case, these are policies presented by central government as its policies; in the formation of such policies groups of members of Parliament, or even individual members, may exercise some influence, but it is clear that in general, Parliament today has a less significant influence in the formation of policy than do a variety of other bodies, from the political parties to the trade unions, employers' organizations, and other major interest groups. Occasionally (though with great rarity in the field of economic policy) an interest group may short-circuit the normal process of policy formation and implementation, addressing itself directly to Parliament, via sympathetic members, for the enactment of legal measures to secure its policy objectives. Such measures are too rare (and given the present balance of constitutional powers in the United Kingdom, unobtainable in this way unless the government itself adopts an attitude of neutrality or tolerance) to warrant treating Parliament's role in economic policy as other than an essentially reactive one.[18]

The same cannot be said about the courts. Courts do, after all, take decisions with binding effect, not only for the individual parties before them but also, through the operation of doctrines of judicial authority and precedent, for all parties who now or in the future find themselves in the

same situation. In the civil sphere, the decisions are usally reached without the help or intervention of any other organ of government. Such decisions might, in my terminology, be measures, if in arriving at them courts are seeking, consciously or even unconsciously, to achieve goals in the nature of economic objectives. In developing the common law of restraint of trade, for example, the courts may be seen as attempting to inhibit, through a judicial policy of non-enforcement, the use of contractual devices to create or consolidate dominant positions in local and other markets.[19] More fundamentally, the whole activity of the courts in developing and maintaining the principles of the common law is viewed, by the economic analysts of the Chicago School,[20] as promoting a basic economic objective of rational allocation of resources for which all policy-makers should strive. Part of the Chicago prescription, in fact, is that courts, in administering the common law of tort and contract, may offer better economic policy than do governments and legislatures engaged in market-distorting regulation.[21] These considerations raise the question whether we should therefore regard the common law, in the hands of the courts, as a source of economic policy independent of the actions of central government.

The answer is a qualified "no". What the restraint of trade example and the work of the economic analysts of law demonstrate is that the common law has an economic content, and that rules of law shape economic relationships. As I have already argued, the essentials of the common law – rights of personal security and property, for example – should be viewed not as part of economic policy itself, but as a framework within which that policy is contained and against which it may react.

Some parts of the common law, however, can be seen changing and developing: the restraint of trade doctrine, already cited, is an obvious example. Here the courts are not affirming or reinforcing a fundamental principle but are setting the appropriate scope of an exception to such a principle,[22] a task which may be viewed as falling within the ambit of the activity of economic policy. Even here, however, the process of change has generally been slow, and the courts have echoed policy developments that were occurring elsewhere, rather than themselves leading the way. Thus the vigour of judicial application of the restraint of trade doctrine has waxed, and waned, and waxed again, responding first to legislative condemnation of monopolies and combinations in Tudor and Stuart times, then, as already noted, to *laisser faire* ideas and the repeal of restrictive legislation, and now to the rebuilding of a structure of competition legislation since the Second World War. There have, it is true, been time-lags in this process, where slowness of judicial adaptation to legislative policy has resulted in a distinctive judicial attitude, as in relation to combinations of workmen at the end of the nineteenth century.[23] There may also be abrupt changes of direction, as the judiciary attempts to

follow, in its common law doctrine, the sinuous and ill-marked track of executive and legislative policy: a recent example is afforded by the apparent espousal, by the House of Lords, of inequality of bargaining power as a ground for invoking the restraint of trade doctrine.[24] At such times, perhaps, the courts may be regarded as an independent source of economic policy; but the scope left by legislation for the development of the common law in areas relevant to economic policy objectives is so small that the courts' importance, in this role, is minimal. This is *not* to say, of course, that their importance in relation to economic policy is minimal: their approach to, and their decisions in, their tasks of interpreting, reviewing and enforcing, legal measures of economic policy remain matters of central concerns.

## The relationship between objectives and instruments

### 1. Quantifiability, aggregation, and modes of action

For Kirschen, it will be recalled, an economic objective was a concept which could be given "some quantification".[25] Quantification of objectives is both an essential tool of processes of economic modelling and forecasting, and a largely accurate description of how government assesses its performance in the economic sphere. Progress in full employment policy, for example, may be measured in terms of the ratio of unemployed workers to job vacancies, or of unemployed workers as a percentage of the total labour force. Progress towards price stability may be measured in terms of the annual increase in the retail price index, or some more broadly based cost-of-living index; progress in regional policy in terms of ratio of income per head between the poorest region and the national average. Some objectives, such as an increased rate of technological innovation in industry, or a greater propensity to competitive behaviour, are harder to quantify, and the measures chosen (in these cases, for illustration, numbers of patents taken out annually; industrial concentration ratios) may be open to a variety of interpretations. In all areas of economic policy, though, the relevance and significance of the government's chosen tools of measurement tend to be matters of professional and political controversy. But the general system of measurement and appraisal, involving quantities, targets and results, is taken for granted in the economic policy field, to the point where the numbers may seem to take on an importance of their own, to the detriment of an appreciation of the more complex realities they conveniently represent.

Reliance on quantitative measurement is not, of course, peculiar to economic policy. Success in many other fields is increasingly measured by

reference to convenient statistics: the infant mortality rate; the number of persons undergoing higher education; the size of the armed forces or the time within which combat units can be put in place; the length of the motorway network. A difference of degree should however be noted. It is usually meaningless to speak of success in economic policy otherwise than in quantitative terms, whereas in other areas, though there is a continuing search for new and more subtle quantitative indicators, there remains some recognition that not every aspect of human happiness lends itself to numerical measurement. Economic policy, of course, addresses precisely those aspects that have this quality.

Without too much distortion, therefore, one can view economic policy as the attempt, by government, to influence the movement of certain economic quantities or indicators, whether by restraining movement away from specified targets or promoting movement towards them. The means at the government's disposal for exercising such influence are almost entirely indirect, that is to say, they operate on the actions and decisions of persons outside the government, whose aggregated results, determine the level of the relevant economic indicators. The popular vocabulary of economic management does not reflect this quality of indirectness: we speak of government "creating jobs", "restricting imports", "boosting investment", all phrases which suggest the capacity of government, of and by itself, to secure the results referred to. We do not, however, live under the kind of centrally planned economy within which the policy-maker has direct control over the factors of production and can issue orders for large-scale changes, in, for example, the balance of production as between consumer and investment goods. Even in developed socialist States such as the USSR and other countries of Eastern Europe, the reality of the planned economy is far more complex and uncertain than the simple model of plan and command would suggest; and in the United Kingdom, as in other Western "mixed" economies, it is only in the public sector – indeed, only within that part of it directly controlled by central government – that the possibility exists of altering the quantities expressive of economic policy objectives by direct executive decision.

Notwithstanding the present size of the public sector in the United Kingdom – since 1974 public expenditure has varied between 40 and 47% of gross domestic product[26] – the scope for such direct executive decision is limited. Central government expenditure, which runs at about 66-69% of the total, is heavily concentrated in the areas of defence and social security, and only a minor proportion (about 40%) goes on the purchase of goods and services (including labour), where direct effects may most readily be produced.[27] The rest is spent on subsidies, grants and loans to individuals, enterprises and other public authorities. While this 40% is still a tidy sum in money terms – roughly £25 513 million in 1979-80 – its redeployment in the extent necessary to produce significant changes in

any major economic aggregate could have major adverse effects – at least in the short term – on the discharge of public functions.

Unless, therefore, the government is hypocritical and its policy is hollow, in the sense that it adopts as an objective a quantity which it thinks would occur anyway, without intervention (a possibility not to be excluded as unthinkable), the setting of a policy objective implies setting out to affect the behaviour of others: the "others", in the public sector no less than in the private, who charge prices, pay wages, export and import goods, invest capital, borrow and lend money, make take-over bids, purchase goods and services. It is their actions which in aggregate or on average make up the greatest part of all the quantities which government is trying to manipulate; their actions, therefore, which must be made different from what they would otherwise, in the absence of the policy, have been. More precisely, *some* of those actions, *some* of that behaviour, must be different: zero and 100% are not necessarily the only quantities that government aims at. A government that wants a rising birthrate for economic reasons may not wish every wife to bear an extra child. It is important to understand that though government is interested only in totals or averages, the relevant actions cannot be manipulated in the mass; each separate economic actor must be addressed, but only with a view to getting a sufficient number to change their course of action to satisfy the demands of the policy. Aggregated ends are pursued by individualized means.

## 2. Television imports: a hypothetical case

These propositions may be clarified by a hypothetical example. Suppose that an object of governmental concern is the viability of the domestic television manufacturing industry; that a situation has arisen in which imports are capturing an ever-increasing share of the home market. Foreign sets are, it appears, more reliable and technically more advanced than their home-made counterparts. The domestic industry represents to government that given time and opportunity to invest in new manufacturing facilities and equipment, it could be fully competitive and could recapture and retain an adequate market share. The need, therefore, is to provide a breathing-space by securing that for a certain period, imports of foreign sets are held at the current level or, better, reduced by some indefinite amount. Let us assume that the government, worried about deindustrialization and unemployment generally, accepts this argument for temporary protection. It confronts a market in which two chains of supply – on the one hand that of foreign manufacturers, their importers and distributors, on the other that of domestic manufacturers and their distributors – merge at the retail level where sets are sold to their ultimate users. Even

with the specific objective mentioned, this situation offers a wide choice of points and modes of intervention.[28]

The government may decide to address itself directly to those who are exercising choices as between domestic and foreign sets, that is to say, consumers and retailers. It may do this by means of a "Buy British" campaign of ministerial speeches and public advertisements; by offering a financial inducement to purchase a British set, such as a reduced licence fee; by imposing a financial penalty on the purchase of a foreign set, such as a point-of-sale tax, purchase tax, or discriminatory rate of VAT; or, most forcefully, by securing legislation to make it an offence to buy, or as a retailer to sell, a foreign television set. Alternatively it may intervene at an earlier point in the supply chain, as by reducing the costs of British sets by a subsidy to manufacturers, or by increasing those of foreign sets by an import tariff payable by importers, or by the stipulation of unusual manufacturing standards for the British market which foreign manufacturers would find it difficult and expensive to meet. Finally, the government may aim directly for a reduction in supply, by negotiating export restraint with foreign manufacturers or with their governments, or by the temporary prohibition of imports, or the fixing of an import quota.

## 3. Distinctions between instruments

This example may be used to suggest a variety of distinctions which may be drawn among the instruments available in relation to a given policy objective.

(i) We can contrast instruments by reference to their connection (or lack of it) with legal norms. A publicity campaign, or an "understanding" about export levels, seem *prima facie* to have far less to do with law than do prohibitions and quotas.

(ii) Another point of differentiation is by reference to the number of people addressed by each instrument. Generally speaking, the further back up the supply chain the government goes in search of its point of intervention, the fewer the number of individuals whose choices it needs to affect. At the level of purchasers of television sets, numbers may be counted in thousands; at the level of manufacturers or governments, on the fingers of the hand. Interventions at this level are obviously more convenient in administrative terms because of the smaller number of enterprises and transactions on which they need to operate, but this convenience may be offset by the disadvantage of a lesser degree of certainty as to the effects of the instrument. If manufacturers or importers decide to absorb themselves the costs of a tariff on imported television sets, rather than pass it on to the ultimate purchaser in the form of increased prices, the desired effect of reducing import penetration may be greatly attenuated. Whatever

the merits or demerits, from the efficiency point of view, of interventions at different points of the supply chain, it should be obvious that instruments which address large numbers of people may call for different techniques, and hence different kinds of measures, from those appropriate where only small numbers are involved.

(iii) The example also illustrates the point that the government may not be aiming to alter all relevant choices. Its aim, in this respect, will depend, once more, upon its chosen point and method of intervention. If the choice falls on a reduced licence fee for all who purchase a British set within a given period, the government may be taken not only to have accepted the fact that some who buy British and obtain the reduction would have bought British anyway (so that the measure's only effect in this respect may have been an acceleration of the purchase) but also to have calculated that not all who would have bought an imported set will be led to change their mind: the aim of the policy is that a sufficient number should do so. If on the other hand the chosen weapon is an import quota, operated by awarding to existing importers individual licences for a fixed fraction of the quantities previously imported, then it is of the essence that every individual addressed should alter his import decisions in full accordance with the stipulations of the quota.[29] Precision of effect, in this latter case, is purchased at the cost of administrative complexity and the ossification of the market structure, a price which may, or may not, be too high.

## 4. The concept of relative cost: the principle

While these distinctions between instruments are important, more important still is the fact that the instruments illustrated all operate with the same basic purpose and effect. In each case, the aim of government is to change the behaviour of individuals by altering the relative costs of behaviour which is, and is not, in accordance with its policy. This may be done either by reducing the costs of behaviour in accordance with the policy; or by increasing the costs of behaviour which is contrary to it. The costs involved may be money costs; but they need not necessarily be so. Quite obviously, a discriminatory sales tax will make an imported set relatively more costly than its untaxed domestic rival; a discriminatory licence fee reduction will make a domestic set less costly to own. But a "Buy British" campaign, without any deployment of taxes or subsidies, may also make the domestic product relatively less costly, in that for the same money the consumer now gets not only a television set but also the satisfactions of expressing patriotism and solidarity, satisfactions which, though not marketable, are none the less real and valuable. The same analysis may be applied to measures which involve the creation of criminal penalties. The relative cost of the imported set whose purchase is prohibit-

ed on pain of a fine is increased – doubtless in a different amount for each potential purchaser – by the amount of the likely fine together with the costs, financial, psychological and social, of undergoing prosecution, all discounted by reference to the degree of probability of detection and prosecution; and to this must be added a further increment representing the "moral cost" to any particular individual of breaking the law, as this affects his personal integrity and self-esteem, as opposed to threatening him with external consequences. A comparable calculus of costs and benefits is presented to the potential importer of foreign sets when faced with an import prohibition. He, however, may also take into account, on the benefit side of the balance, the increase in the value of the imported set, deriving from its scarcity, which may enable him to recoup in money the costs associated with risks of prosecution.

The value of the idea of relative costs is that it provides a unifying framework within which we can discuss a variety of instruments of policy which may, as already noted, "invoke the law" in different degrees and different ways, or not at all. It offers us a common point of reference for the analysis and comparison of such diverse legal measures as criminal prohibitions, licensing schemes, expenditure legislation and public contracts, and may permit us to develop hypotheses to explain, within the framework of a given economic order and constitutional-legal system, the choice of legal measures in relation to particular policy objectives and instruments, as well as the mode of enforcement of such measures.

## 5. Relative costs: some clarifications

### (i) *Individuals*

When referring in general terms to the relative costs concept I have spoken of affecting "individuals'" decisions, choices, or behaviour. As the television imports example suggests, this term is to be understood as including a wide variety of economic actors: not only real people, acting singly as consumers, employees, and so on, but also enterprises, decentralized units of government, even foreign governments. "Individual", therefore, is used as a shorthand description for the distinct decision-taking unit whose choices government seeks to affect, whatever may be its nature or size, and regardless of its legal status. As used here it connotes neither the singular physical person nor the incorporated legal person: it refers uniquely to the capacity, as a matter of fact, to take decisions and follow courses of behaviour which are of concern to government in relation to its economic policy objectives. An unincorporated association that possesses this capacity – a trade union, for example – is therefore for us an individual, no less than is the consumer purchasing a television or the

company manufacturing it. Naturally we cannot be indifferent to the colossal disparities – of numbers, size, wealth, power and organization – existing as between members of this broad class. One object of any legal discussion of economic policy must be to examine what distinctions have been drawn, in the framing and application of legal measures, between different types of "individuals", and with what effect.

Nor should we neglect the fact that decision-taking units involving more than one person are not monolithic: decisions are the product of a structure of power within the unit. That structure may be highly complex, involving both the aggregation of the preferences of individual people into group viewpoints of interests, and the balancing of those viewpoints or the resolution of conflict between them. Even within a medium-sized manufacturing company, for example, there are likely to exist distinct groups of workers (divided by factors such as trade union allegiance, possession of skills, mode of remuneration) and while each of these groups will have an interest in the survival and prosperity of the company they may have quite different short-term aims in specific situations such as expansion, redundancy or takeover. In relation to many areas of policy, and to the use of most kinds of instruments, these structures, and the diverse interests they encompass can be viewed purely from the standpoint of their decision-taking capacity as collective entities, which is their passport to "individual" status. In two situations, however, an analysis of policy requires investigation of the structure as such. The first case is where the power-relations within the decision-taking unit affect the way in which the government approaches the task of affecting its decisions. Recent British Government policy towards the National Coal Board, which has been largely conditioned by the acceptability of decisions to the Board's employees, affords a good example. The second is where the alteration or creation of structures is itself the instrument, or an instrument, of government policy. New corporate forms as a support for technological innovation in small enterprises, worker participation as an instrument for greater productivity though increased industrial harmony, are but two possible examples from a rich field. In such cases, however, the relevant individuals are those who operate or will operate within the structure, whose relative costs are affected by the changes proposed in their opportunities or constraints.

### (ii) *Control and choice*

Implicit in the relative costs concept is the idea that the individual decision-maker always retains a choice as to whether he will align his conduct with the demands of government policy, no matter what instrument government deploys. It is not difficult to imagine – though much harder actually to find – situations in which the physical control and

supervision exercised by governmental agents is so tight as to eliminate even the possibility of non-compliance, so that choice is absent and non-compliance beyond price. The rarity of such cases, however, serves essentially to emphasize the element of choice existing in all normal cases, even in the face of express prohibitions. The point is worth stressing, not least because there is a tendency among writers who set out to assess the costs and benefits of using different kinds of instruments in fields such as pollution policy, to assume that people always obey mandatory legal rules.[30] On this basis regulatory standards are argued to be inflexible and productive of sub-optimal results, in contrast to "market-type" instruments such as taxes, subsidies, or tradeable pollution entitlements. These are said to leave sufficient discretion to the individual to permit him to adjust his activity in a way which is capable of achieving the best available balance of compliance costs and policy benefits. Behaviour in response to mandatory rules is in fact much more complex than this model allows for: in the economic sphere, at least, calculated and negotiated non-compliance are common phenomena, and are based on the same kind of cost-benefit analysis as is explicitly demanded by the use of "market-type" instruments.[31] There may still be very good reasons for preferring, in a given case, a tax-based to a regulation-based scheme (for example, greater economic transparency or the reduction of administrative discretion): but the evaluation must take account of the individual's "discretion to disobey".[32]

Even the use of criminal sanctions to support prohibitions does not disrupt the essential unity of the relative cost concept. The traditional association of the criminal sanction with moral imperatives has been weakened almost to the point of disappearance in the field of economic policy by this frequent use in situations which imply no moral stigma, as is indicated by the fact that governments often interchange economic instruments which involve reliance on criminal penalties with others that do not – for example, substituting import duties for import quotas, and vice versa. In the context of an inquiry into the deployment of legal measures in the service of economic policy objectives, any residual moral implications attaching to criminal prohibitions are important only in so far as they affect the processes of choosing and operating legal measures. It is conceivable that in a particular case the moral implications associated with the criminal sanction may make a criminal measure inappropriate, or, once enacted, may inhibit its application; conversely the presence of such a sanction may in some cases reinforce the efficacy of a measure by adding specific moral costs to detection and conviction. The fact that the criminal sanction may, in certain applications, possess these distinctive qualities in no sense requires its separate treatment, but reinforces the need for a common concept to which these distinctive qualities, along with those of other kinds of measure, may be related.

### (iii) *Knowledge and experimentation*

Government, it may be said, can surely never know the individual economic agent's calculation of the costs and benefits of a particular decision. In such a state of ignorance, can it make sense for government to set out to alter such a decision by adding in futher costs and benefits whose value to the individual it likewise cannot know? In most cases, however, such knowledge is not vital. What is important to government is not any specific individual decision, but the economic aggregate, the sum total of such decisions. It may well not care which individuals change their decisions in response to its policy, so long as a sufficient number of them do so. Indeed, the policy may not even aim at any specified quantitative effect on the relevant aggregate, but may simply seek to increase or reduce it. In the latter case, the imposition of tax of an arbitrary amount will produce an observable reduction, changing the decisions of those agents for whom the calculated benefit from the proposed decision was less than the amount of the tax. Assuming that it can get accurate information about such aggregate changes, government can then use this information to refine its policy and improve its powers of predicting the results of future tax changes. The technique is essentially experimental in character and does not depend for its success on knowledge of the decision-making of individuals. It none the less remains the case that the measures actually adopted produce their effects through the process of altering individual decisions already described. To avoid confusion, it is worth reiterating that this experimental approach is not the only one which may be pursued by government. By intervening at a point, and in a manner, which enables it to deal with smaller numbers, government may put itself in a position to acquire the information needed for the refinement of policy by other means, such as negotiation, bilateral discussion, and so on. At the same time it may seek, as in the example of import quotas, to change the behaviour of this smaller number of individuals in a more precise way.

## 6. Choice of instruments: an hypothesis

I said earlier that government may seek to change behaviour in two ways: by reducing the costs of behaviour which is in accordance with its policy; or by increasing the costs of behaviour which is contrary to it. These two ways of proceeding may come to much the same thing, if the individual has before him only two choices wich are relevant to the policy. Say the government decides to resolve its television imports problem by approaching the consumer directly. Since it is reasonable to suppose that a much higher degree of substitutability exists between domestic and imported television sets than between either kind of set and other things, govern-

ment can confine its interest to those consumers who are going to buy a television set. These people will be confronted with a wide range of choice (colour/black and white, large screen/small screen, portable/fixed, etc.), but their only choice that matters to government is between foreign and domestic makes.[33] There being only, for this particular set of consumers, two possible choices, an increase in the costs of one is equivalent to a reduction in the costs of the other. Government may therefore produce the same kind of result by either of the two ways of proceeding (though a given cost reduction will not give exactly the same quantitative result as the same cost increase).[34]

The picture changes when the relevant individual choices cannot be reduced to two. If government takes the view that the only available solution to the problems of British television manufacturers is modernization of plant to secure lower-cost or higher-quality production, and takes modernization as its policy objective, its obvious course is to reduce the costs, to manufacturers, of investment programmes for modernization. The alternative is to try to design a cost-increasing measure or measures that would bear upon all choices as to use of funds alternative to investment for modernization, choices which might range from increased dividend payments, through increased wage and salary scales, to such sumptuary expenditures as increases in the number of executive washrooms or of the chairman's *poules de luxe*. The only measure which would cover and penalize all such choices would be a requirement, backed by civil and criminal penalties, that certain funds of the manufacturers be applied to such investment programmes. This would cost the government less, in money terms, than an investment subsidy.

The adoption of such a measure in normal circumstances is hard to imagine, by reason of its incompatibility with prevailing ideas about the general relationship between government and private individual and corporate property-holders. In terms of the present analysis, such a choice of instrument is, in the United Kingdom, practically (though not legally) precluded either because it offends against values engendered by the pursuit of non-economic aims, or because it falls outside the space set for "process policies" by the national economic order, or both.[34a]

The examples in the two previous paragraphs suggest a general hypothesis which can be tested in any detailed review of economic policy implementation. It is this: the more precise the policy objectives government sets itself, the more likely it is to see an advantage in proceeding by way of reducing the costs of desired choices, rather than increasing the costs of undesired choices.[35] The more precise the objective, the narrower the segment of the range of individual choices that are consistent with it, and the wider the segment of inconsistent choices. To revert for a moment to the first example in this section, if government's only preference is for consumers to buy domestic, rather than imported television sets, the

cost-reducing and the cost-increasing approaches may do the job equally well; government's main concern may be with the costs – financial, social, political, administrative – that it incurs in carrying either type of approach into effect. But if government's preference is for consumers to buy in substantially larger numbers a range of colour sets made by a particular British manufacturer, the design of a measure to increase the costs of all available alternatives is all but impossible, and a cost-reducing measure (for example, a subsidy on each such set) will be preferred.

## The nature of policy instruments

### 1. Power-resources

If government is to change the relative costs attached to individual behaviour, it must be in a position to deploy certain resources. In this it does not differ from any person who seeks to influence the actions of others: that person, if his attempts are not to be vain, must have available some store of sanctions and inducements, some source of power. The resources that constitute power will vary according to the nature of the relationship within which that power is to be exercised: power in the family may be far more multi-faceted than power in national politics. Tears are a potent weapon between lovers; they have their uses in the courtroom; but as the sad case of Lord Lundy showed,[36] they melt no hearts in Parliament. Within the framework of economic policy, we are concerned largely, if not exclusively, with impersonal relationships, where people interact as representatives of institutions (for example civil servants, company executives), or in the performance of given functions (for example, judges), or for the purpose of private economic interests (for example, consumers and retailers). In this context natural affection is of little consequence, and the resources we are concerned with are, it is suggested, essentially three: force, wealth and respect.[37] When government forbids the purchase or importation of foreign television sets, or taxes their importation or sale, it employs (if only by way of threat) force; when it subsidizes domestic manufacturers or consumers it employs wealth; when it urges us to buy British it invokes respect – for itself as the duly elected and constituted holder of economic responsibilities. There are of course interconnections between these three resources. Particular instruments may rely on an inextricable combination of two or more of them: voluntary import restraint agreements with foreign governments perhaps provide an example. In the last resort, the government's possession of force is dependent both on its use of wealth to pay its soldiers and police and their respect for its authority. The wealth of government, in turn, stems largely from its predecessors' use of force to acquire and subjugate territory and from its own continuing threat of force in the raising of taxation.

These interconnections may be peculiar to government, but the possession of these resources, in some form and in some measure, is not. All but the very weakest can deploy force in at least some personal interactions, and in the economic sphere, such force – or the threat of it – is regularly employed by private individuals to affect the choices of others. Examples include armed robbery, the patrolling of land by armed keepers or guard dogs, and the operation of protection rackets. A high degree of sophistication and systematization in the employment or private force may be attained, as exemplified by the activities of the Mafia in Sicily and southern Italy. The employment of wealth by individuals in pursuance of economic objectives is of the essence of everyday life: every consumer transaction can be analysed, both from the viewpoint of the buyer and of the seller, as the use of wealth (on one side in the form of money, on the other in the form of goods or services) to alter the decision of the other party from that of retaining to that of transferring what he holds. Bribery and corruption exemplify other uses of this resource. Examples of the private possession of respect are perhaps harder to find in the economic sphere (there must be some form of non-governmental authority that is the subject of respect, such as the authority possessed by the church in the moral and religious sphere), but we may perhaps see its deployment – not always with great effect – in the publication of economic forecasts and commentary by private institutions.

## 2. Private power-resources and the law

The treatment in the legal order to these private resources of power is fairly complex, and varies substantially as between one national legal order and another. In the United Kingdom, the law closely restricts the private use of force, permitting it only for the protection of person and property, and then subject to fairly stringent limitations. While legitimate private force is therefore exceptional, the legitimacy of the deployment of private wealth remains, in the United Kingdom, the rule rather than the exception. Certainly, specific limitations on the use of wealth become more numerous almost day by day: entire bodies of law with this function – planning law, environmental law, consumer law – have grown up in the present century, adding to older prohibitions, as of bribery, which attacked different evils. Though there may come a point where this multiplicity of singular restrictions combine to cut the heart out of the right of property, that point has not yet been reached in the United Kingdom: money, or money's worth, still buys most things somehow. Moreover, it is the law which, through the institutions of property, contract and inheritance stabilizes the possession of wealth and enables it to be used as a resource. As to respect, this is a resource with which the law has always

found it hard to grapple, doubtless because its working, unlike that of the other resources, is seldom manifest to the senses. In pluralist and libertarian societies like that of the United Kingdom, attempts at its control by law are, in addition, likely to be controversial insofar as they involve the control of speech and opinion.

## 3. Governmental power-resources and the law

### (i) *Force*

While government must share with private individuals the resources of force, wealth and respect, it does not suffer the same legal restrictions on their use. Clearly, government is in a privileged position when it comes to the lawful use of force. Seen as a temporal process, the restriction of private force is but the corollary of the development of a near-monopoly of force in the hands of government, as self-help and private revenge were replaced by public enforcement, and private, local, merchant and ecclesiastical jurisdictions were all gradually mastered and dismembered by the king's law. This was not, however, a process of concentration of unfettered physical force. What was being accumulated in the king's hands was power subject to law and for the enforcement of law, and while the steady destitution of other centres of physical force made the king harder and harder to challenge on the basis of might, there was present at almost all times the conception that his power was to be exercised according to right. The determination of what was a right exercise of the king's undoubted power was a matter for law. The issue, in the legal, constitutional, and physical battles waged through the seventeenth century in Britain, was not whether there were legal restraints on the king's power – none sought to deny that – but whose law was to furnish those restraints, who was to make, interpret and apply that law – the king, his administration, and its internal courts? or parliament and the common law courts?[38]

Few will need to be reminded that the result of that struggle was a virtually complete victory for the interest of Parliament and the common lawyers, as expressed in the Revolution Settlement of 1689 and its constitutional documents, the Bill of Rights in England and the Claim of Right in Scotland. What this signified in terms of the argument here was that the wishes of the Crown, so far as they depended for their sanction on the deployment of force, could no longer be made binding on the subject save through the device of parliamentary legislation.[39] The King, in the person of James I and VI, had been told by the courts as early as 1611 that he could not impose by proclamation even such sensible policies as the restriction of new building in London or the prohibition of the making of starch from wheat: for such regulations to be legally effective they must be

embodied in Parliamentary legislation.[40] This had not prevented frequent recourse to proclamations by later Stuarts,[41] and their vigorous enforcement through the prerogative courts, principally the Star Chamber in England and the Privy Council in Scotland; but with the disappearance of such courts this disputed instrument of policy was finally wrested from the hands of the Crown.[42] Today, therefore, legislative authority, direct or delegated, is needed not only to create criminal prohibitions and other legal restrictions on behaviour and to impose taxes,[43] but also (since the ultimate sanctions of private rights are remedies which rely on public enforcement) to alter legal relationships between private individuals. Remarkably enough, there is still room for dispute, three centuries after Parliament purportedly abolished the royal suspending and dispensing powers,[44] as to whether legislative authority is likewise necessary for the relaxation or systematic non-enforcement, or partial enforcement, of such restrictions, or the non-collection of taxes.[45] Apart from this exception, and the preservation of certain limited prerogative powers,[46] we may categorically assert, as a constitutional principle, that policy instruments involving the deployment of force by government require legislative expression or authorization.

### (ii) *Wealth*

It is hard to maintain such a clear view when we approach the second power-resource in the hands of government, wealth. For a start, government, in the United Kingdom as in other Western States, enjoys no monopoly of wealth or of its lawful use. Even when we look at particular kinds of wealth, at the legal regime relating to the possession of particular resources, we find that relatively few are reserved for the exclusive use and enjoyment of the State. With the exception of certain minerals – gold and silver,[47] petroleum,[48] the mineral resources of the Continental Shelf[49] – most such reserved resources are closely associated with the government's monopoly in the legal use of force: military equipment is the obvious example.[50] So long, however, as the possession and use of most forms of wealth is considered, by a society, to be sufficiently benign to warrant leaving such power in private hands, it is unlikely that we shall see the kind of determined quest for its control and restriction when in the hands of government that characterized the approach of 17th-century parliamentarians and lawyers to the royal monopoly of force.

At one time we might, indeed, have expected to find the possession of vast wealth viewed as a positive attribute in a monarch, in so far as it was likely to restrict his resort to taxation for the financing of royal activities. If this now seems an anachronistic viewpoint, it is because over the centuries there has always been a gap – and it has been an ever-widening

gap – between the costs of the activities undertaken by government and the capacity of government to defray those costs from the fruits of its "own" wealth. The demand of kings and governments for aids from the population in the form of taxes, intermittent and of varying intensity up to the seventeenth century, has been permanent and of increasing insistence ever since. It lies at the root of the development of a specialized legal regime to regulate government's deployment of its wealth, a regime which, with one possible exception,[51] has a quite different historical finality from that which, as we have seen, gives expression and authorization to governmental deployment of force.

One of Parliament's main concerns, in the period when it was the occasional recipient of requests for aids from the Crown, usually for the purpose of defraying the costs of military adventures, was to ensure that the funds raised by way of the taxation it consented to were in fact expended for the purposes for which they had been demanded. Such a control of the application of the funds might reduce, though it could never eliminate, the risk that the king would shortly return, pockets empty, with renewed requests for the same objectives. Thus there were, in the 14th century, occasional parliamentary attempts to appropriate taxes to defined uses and to secure an accounting for expenditure.[52] In the Tudor period Parliament was too weak to achieve any success in this direction. As the expenses of government under the Stuarts outran ever further the wealth of the Crown, and their defrayment from taxation became continuous and generalized, so too did the concern of Parliament; the desire to ensure that specific taxes were applied to the specified ends became the desire to see a regular and provident deployment of the funds raised from general taxation. Little progress in this sense was made during the Stuart period itself,[53] but the attachment of appropriations to tax legislation became regular practice after 1688, and gradually developed into a process of annual appropriation to specified purposes of all revenues enjoyed during the year, a process which arrived at something like its present form by the beginning of the nineteenth century.[54] By reason, no doubt, of its origins in the earmarking of specific taxes, themselves authorized under the forms of legislation, the act of appropriation of funds by Parliament has always taken legislative form, though there exists no judicial or other constitutional declaration, comparable to the *Case of Proclamations*,[55] to the effect that government spending must be covered by such legislative appropriation.[56] In fact the practice and procedures of annual legislative appropriation are now too well established for the requirement to be seriously called in question, though there remains ample room for argument on the related question of when some permanent and detailed, as opposed to annual and outline, legislative authorization of government expenditures is constitutionally requisite.[57]

While, therefore, we may say that it is today common-place and normal

to find that the deployment of wealth by government is authorized in some form by legislation, we cannot be as categorical as in relation to the deployment of force. Nor can we assume that legislation serves the same function in each case. Certainly there is a common factor, in that the legislative process secures the civic benefits of comprehensive discussion, democratic consent, publicity and formal promulgation for governmental measures. It controls, in short, the deployment by government of its power-resources; though in these years of stable parliamentary majorities and comprehensive governmental concern for all aspects of public and private welfare, the government so continuously seeks the enactment of legislation, and obtains it with such facility, that this control function is easy to forget. But the historical reason for invoking control is different in the two cases. Force is seen as dangerous in itself, and the purpose of the constitutional requirement of legislative consent is to protect the subject against the oppressive ruler. Wealth, however, with the exception of its use to maintain armies, always the subject of a specially strict Parliamentary scrutiny,[58] is seen as in principle benign, and the purpose of requiring legislative consent is not to protect the subject against its oppressive use but to protect the collective interests of the taxpaying public against its improvident use.

The contrast here drawn, which is central to my analysis of law in relation to policy, is between the different bases of the constitutional requirement to legislate. It is not concerned with the reasons for which the government may seek to persuade Parliament to pass particular pieces of legislation. Often government will have no choice but to propose legislation if it wishes to employ given instruments of policy, by reason of the operation of the constitutional principles which have just been discussed. Where the government does have a constitutional choice whether to legislate or not (as when it has to decide whether to constitute a new public body as a chartered corporation or limited company, which will not need legislation, or as a statutory public corporation, which will) its decision is more likely to be determined by considerations of administrative and political convenience, or of the importance of publicity, than by any worries about its propensity to act oppressively or improvidently.

Parliament's current perception of the nature of its legislative task cannot, of course, change the historical bases of that task. When constitutional requirements for legislative action were established they represented the means of securing a system of checks and balances as between the executive government and a parliament representative of the propertied and merchant classes whose common law rights (and particularly their rights in property) were most sensitive to government action and who were called upon to defray the major (and always increasing) part of the expenses of the State. But the development of universal suffrage and the rise of disciplined mass parties have changed both the character of

members of the dominant House of Parliament and the factors affecting their parliamentary behaviour. There is now a permanent majority of MPs for whom the issues over which Parliament strove to assert its right of legislative decision are secondary to the question of the effectiveness of government in working toward its various economic and social objectives. Even the minority – mostly on the right wing of the Conservative Party – whose political priorities are closer to those of the 17th and 18th centuries can usually be stifled by the application of party discipline. In consequence there is no longer an inherent reluctance to legislate in aid of public policy, and the control function of the legislative process is thereby substantially weakened. This weakness of operation, however, while it may have a significant impact on the results of the legislative process, cannot affect the reasons for which that process is required to be invoked.

### (iii) *Imperium and dominium*

The difference in the historical reasons for requiring legislative expression of governmental deployment of force on the one hand, and wealth on the other, is reflected in differences – persisting till the present day – in the style and effects of these two kinds of legislation, and in their relationship with rules of the common law. Legislation relating to force focusses on the individual and the way in which his legal situation – his existing legal rights, freedoms, powers and duties – is to be affected. This applies whether the law directly alters that situation, as where a new prohibition is imposed or tax levied, or whether, as is more common, it confers powers on government, in broad or narrow terms, to take action which has such effects. Legislation relating to wealth, in contrast, focuses on the commodity itself: it lays down rules, expressed by reference to purposes, conditions or quantities, to regulate the disbursement of funds or property by government. While it may confer on individuals rights to receive such funds or property in accordance with those rules – social security legislation offers an example – it does not necessarily (indeed, in the sphere of economic policy, normally) do so.[59] In cases of the latter type the terms, and even the nature, of the transaction through which the deployment of wealth produces its effects on the individual may not be specified by the legislation at all, the government being left a wide discretion as to the choice and form of legal means.[60] The contrast between these two ways of legislating is presented in an extreme form by British financial procedure, which splits the tax side of the budget from the spending side. On the spending side the annual Consolidated Fund (Appropriation) Act simply indicates the amounts which may be spent during the financial year by the government, both in aggregate and on each specific function.[61] On the tax side the aggregate sum to be raised is no less important to the government's

calculations, but the Finance Act says nothing about this; it concerns itself entirely with the rules as to the incidence and amount of tax to be paid by individual persons and on individual transactions.[62]

Many other examples could be adduced to illustrate the point made here: that the nature of the power that government deploys shapes the characteristics of the law through which the power is expressed and controlled. The distinction between force and wealth as governmental power-resources is thus of fundamental legal significance, and merits a specific and unambiguous terminology.

The Latin terms *imperium* and *dominium* express, in an approximate way,[63] the contrast I wish to draw here between the deployment of force, which is an inherent component of rule, and the deployment of wealth or property in the hands of the ruler. I therefore use *imperium* as a generic term to describe those instruments of policy which involve the deployment of force by government (recalling here that force usually means the *threat* of force); *dominium* on the other hand describes those policy instruments which involve the deployment of wealth by government. Imperium-law, therefore, is that body of law which authorizes and expresses the use of *imperium*; dominium-law the body of law which authorizes and expresses the use of *dominium*. In each case the principal component of the category is the body of legislation already referred to; but it will be convenient to use imperium-law to refer to those few remaining rules of *common* law – principally concerned with emergency prerogatives and with police powers – which authorize the use of force by government; and to include within dominium-law those legal devices of the common law, such as contracts, gifts and other transfers, through which the wealth of government may be deployed.

Imperium-law and dominium-law are not, it should be stressed, comprehensive categories which can be made to embrace all law or even all legislation. The case above-mentioned in which government has a choice whether to legislate or not provides an illustration. What was in issue there was the establishment of a corporation, an "artificial" legal person. The common law recognised the utility of this device for the permanent holding of collective property, but also recognized its dangers in facilitating the constitution of centres of collective wealth and power which might rival the authority of the king. With certain exceptions based on long usage it therefore required an act of State power for the constitution of such a corporation.[64] The common law mode of satisfying that requirement still subsists in the shape of the Royal Charter, but alongside it there exist also the possibilities of resorting to an individual act of State power in Parliament, for the legislative incorporation of a statutory corporation, or to executive incorporation under general legislative authority, such as occurs whenever a company is registered under the Companies Acts. Here the law's essential function is to remove a common-law disability in regard

to the holding and transmission of wealth, and it cannot be directly related to the exercise of either *dominium* or *imperium*. This is not to say that no links exist. This kind of law provides a framework for the deployment of wealth in private hands,[65] and may perform the same function in relation to *dominium*, by constituting the public corporate bodies – local authorities, public corporations – through which public wealth is to be held or used.

(iv) *Respect*

Just as imperium-law and dominium-law do not exhaust possible legal categories, so too *imperium* and *dominium* do not describe the whole range of power-resources available to government. Earlier I mentioned respect as another power-resource on which government could call. It is in fact a resource on which government frequently draws, particularly when exhorting individuals to accept its diagnoses of the economic situation and to follow the advice it consequently formulates. It cannot, however, be treated in the same systematic way as *imperium* and *dominium*, because of its at best tenuous relationship with legal ordering, already noted.[66] This is not to say that no connections can be traced between law and respect; one or two examples may indicate their nature. Legal rules may have been designed to protect the resource of respect possessed by organs of government, either by protecting the secrecy of their proceedings or by shielding them from criticism. In the United Kingdom rules of the first type are, at least on paper, considerably more restrictive than those of the second : there is a draconian (and largely unworkable) Official Secrets Act, but no offence of insulting the holder of a public office.[67] Legal rules may also be designed to guarantee that the resource is not misused: the common-law rules regarding undue influence have this kind of aim.[68] Comparable legal rules relating specifically to government are hard to find, their place perhaps being taken by the operation of procedures of political accountability. We may note, however, that in several West European States there has been an attempt to guarantee, by the creation of appropriate legal structures, the independence and objectivity of public economic forecasting and advisory agencies.[69] The only small step in this direction taken in the United Kingdom has been the legal requirement that the Treasury make available for public use the model it uses for forecasting the development of the economy.[70]

Apart from these particular cases there is the general point that respect may reinforce the operation of *imperium*. The State threatens its subjects through imperium-law, but in publicly indicating lines of desired or undesired conduct it also exhorts them to compliance with its preferences. Where the threat of the application of force is remote or absent (for example, where the administrative apparatus for the implementation of

the legal rules has not been placed in position) the mobilization of respect may be the only function actually served by the law.

## 4. Legal categories and policy instruments

How do these discussions of legal categories relate to the earlier development of the argument? I have defined economic policy as the activity of pursuing quantitative objectives through the alteration of individual choices. This is done by changing the relative costs of alternative choices in a way which favours the choices which are in line with a given objective. Government has available for this purpose resources of force (*imperium*), wealth (*dominium*) and respect. *Dominium* and *imperium* are each deployed through law and subject to the control of law, and imperium-law and dominium-law have distinct characteristics. This is the point reached so far. To complete the model of the relationship of law with economic policy, the categories of imperium-law and dominium-law must be connected with the central activity of economic policy, the alteration of individual choice.

Relative costs may be altered, I argued, either by reducing the costs of the individual behaviour desired by government or by increasing the costs of behaviour which government regards as undesirable. Ways of increasing costs include taxes, and regulatory instruments like quotas, standards and simple prohibitions, whose breach carries the threat of fines and physical and other punishments. These instruments all fall within the category of *imperium*, in so far as they invoke, directly or at one or more removes, the resources of force that are at the disposition of government. Conversely, the main cost-reducing instrument mentioned earlier, the consumer or manufacturer subsidy, is an example of *dominium*, the deployment of the wealth of government. Clearly, there is a strong association between *imperium* and cost-increasing on the one hand, and between dominium and cost-reduction on the other. But there also exists, at the same time, an inverted pair of relationships, between *imperium* and cost-reduction and between *dominium* and cost-increase. The creation of an exception to a generalized system of regulation or taxation, or the relaxation or even withdrawal of such a system, reduces, relative to the costs imposed under the general or pre-existing system, the costs of engaging in the excepted or "liberated" behaviour. In like fashion, a selective or general discontinuance of a given exercise of *dominium* will increase the relative costs of the erst-while recipients. Industrialists may be encouraged to invest by tax concessions, or relaxation of planning controls, as well as by investment grants; they may be discouraged from discriminatory labour practices by the withdrawal of government contracts no less than by the imposition of fines.

## Conclusions

My primary aim in embarking on this analysis has been to demonstrate, in a systematic way, the main connections between the activity of economic policy and the occurrence of legal change. The resulting model of the law-economic policy relationship is capable, I hope, of a number of practical applications. Here are two possibilities.

### 1. Understanding constitutional "problems"

When in 1975 the government, with a weak parliamentary position, used its power as a major purchaser of goods and services to enforce its incomes policy, political and academic commentators who disapproved of its action found it hard to express their unease more precisely than by use of the term "unconstitutional".[71] Systematic application of the model here presented might make it possible to specify with more precision the sources of this unease. Careful attention to the *objectives* of the government, for example, would show that they were specified with decreasing precision over the period of the policy, and that in the last, most controversial period, sought only that wage settlements should *average* 10% over the year. Aiming at an average in this way implies flexibility (excessively large early settlements would mean that later ones would need to be held well below 10%) and a case-by-case approach, of the sort usually associated with licensing schemes and similar regulatory regimes. It is clear that on reviewing the *instruments* available to it for this purpose, government found that an equal, if not greater, enforcement effect could be achieved by a "negative" employment of *dominium,* through non-allocation of, and restrictive terms in, its grants and contracts, as through a conventional employment of *imperium*. This course of action had the additional advantage that no specific legislation was necessary, dominium-law, as we have seen, being by reason of its historical functions considerably less specific about the production of effects on individuals than is imperium-law. What caused unease and cries of "unconstitutional" was the government's exploitation of this discrepancy, particularly within the framework of a policy which required the taking of discriminatory decisions.

This type of application of the model provides no "solutions" to constitutional "problems": but it may help to organize the debate, by requiring the critics to specify the element, or combination of elements, in this policy process to which they object, and the particular constitutional principles by reference to which those objections are raised. Is "negative" *dominium* "unconstitutional" *per se*?[72] Are there some "constitutional" limitations to the range of instruments on which government can theoretically draw, or on the substitutability of instruments within that range? Or does the problem really stem from the adoption of a constitutionally

improper *objective*, that is to say, one apt to lead to discriminations in treatment dependent only on the timing of the individual behaviour addressed? Or is it only the combination of these elements in the particular case that is constitutionally problematical? Answers to such questions, in this and other cases, should provide the starting point for discussions of desirable legal or institutional developments which could better protect constitutional principles while respecting, so far as possible, the government's economic policy choices.

## 2. Assessing trends in economic policy

In an influential article,[73] Jack Winkler argues that there exists a pattern of change in the deployment of economic policy instruments, according to which the government is moving from a supportive role and a concern with the general conditions of the economy to a more directive role, in which it seeks to operate directly on the income position of given groups in society. For him one legal manifestation of this process is the use of more direct and specific, but at the same time more discretionary and discriminatory, legal powers. As will be seen, some recent events which have been of concern to constitutional lawyers – such as the incomes policy experience analysed above – also figure in the pattern he observes. Other commentators, looking primarily at different economies within Western Europe, use different terms in describing trends in the methods of economic policy. For the French commercial lawyer Farjat, for example, the important trend is one towards concentration in Western capitalist economies, and the instrumental apparatus of the State is changing in response to this phenomenon. Interventions can, indeed must, be more individualized because economic power is concentrated in fewer centres.[74] Some German lawyers, by contrast, claiming to perceive a general diminution of the scope and strength of formal legal ordering within society, identify a related process which they call "neo-corporatist proceduralism". This describes a tendency for the State to seek procedural, rather than substantive, solutions to economic problems, contenting itself with establishing machinery within which conflicting economic interests or pressures can be associated in a constructive and balanced way, and eschewing the setting of specific targets, for whose achievement those interests must be somehow subordinated and those pressures controlled.[75]

The analysis here put forward is unlikely to enable us to choose between such theories. It can, however, permit their restatement in a common vocabulary through which we can appreciate their common and diverse elements. Even with such a common vocabulary as is afforded by this analysis, such comparisons require caution in that the events said to evidence the trends identified occur in different countries with different legal systems. There seems no reason to suppose, however, that the model

here presented could not be developed so as to be capable of application to the law-economic policy relationship in most Western European legal systems. At that stage, even cross-national comparisons of trends in the legal implementation of economic policy should become feasible.

## Notes

[1] De Smith, *Constitutional and Administrative Law* (3rd ed. 1977) at 208. The sentence has disappeared from the 4th edition (ed., H. Street and R. Brazier, 1981), along with discussion of the developments it referred to.

[2] For fuller discussion see Elliot, *The Role of Law in Centro-Local Relations* (1981), ch. 1.

[3] Wade (H.R.W.), *Administrative Law* (4th ed:. 1977), at 5.

[4] De Smith, *Judicial Review of Administrative Action* (4th ed., 198C, by J.M. Evans), at 15; Wade, *Constitutional Fundamentals* (1980), at 55-57. For more detail on this issue see Ganz, comment (1978) *Public Law* 333; Ferguson and Page, "Pay Restraient: The Legal Constraints" (1978) 128 New L.J. 515.

[5] For a concise and realistic discussion along these lines see Peacock, *The Economic Analysis of Government and Related Themes* (1979), ch. 1.

[6] See in particular his *On the Theory of Economic Policy* (1952), and *Economic Policy: Principles and Design* (rev. ed., 1967).

[7] E.g., Lindblom, "Tinbergen on economic analysis", (1958) 66 *J. Pol. Econ.* 531).

[8] See, for example, Theil, *Economic Forecasts and Policy* (2nd ed., 1961) and other volumes in the "Contributions to economic analysis" series (North Holland, Amsterdam).

[9] See Kirschen et al., *Economic Policy in Our Time* (1964, 3 vols.); Kirschen, ed., *Economic Policis Compared: West and East* (1974, 2 vols) hereinafter referred to respectively as "Kirschen (1964)" and "Kirschen (1974)".

[10] The following account is drawn from Kirschen (1964), Vol. 1, ch. 1, except as otherwise indicated.

[11] Kirschen (1974), Vol. 1, at 17.

[12] For general discussion of the specification of economic policy objectives see Hutchinson, *Positive Economics and Policy objectives* (1964), ch. 4.

[13] Some such specific objectives (for example security of supply in the field of energy) would apparently be treated by Kirschen only as "quasi-objectives" on the ground that though pursued by policy-markes, they have "no welfare content in themselves": Kirschen (1974), Vol. 1, at 18, 23-26. This distinction is without importance for the present analysis.

[14] For a general discussion see Pütz, "Zur Typologie wirtschafts-politischer Systeme", (1964) 15 *Jahrbuch für Sozialwissenschaft* 131, espo. at 141; and for references Schiller, "Wirtschaftspolitik", in *Handwörterbuch der Sozialwissenschaften*, (1965), Vol. XII, at 213-214.

[15] See Robbins, *The Theory of Economic Policy in English Classical Political Economy* (2nd ed., 1978), at. 186-194.

[16] Tinbergen, *Economic Policy: Principles and Design* esp. at. 149, 186; Kirschen (1964), Vol. 1, ch. 6 and at 15.

[17] No special procedures were constitutionally necessary, for example, to effect the major programme of nationalization carried through in Britain between 1945 and 1951.

[18] The same conclusion is reached by the British contributors to Coombes and Walkland, eds., *Parliaments and Economies* (1980), esp. at 91-95.

[19] For an acount of the common law of restraint of trade see heydon, *The Restraint of Trade Doctrine* (1971); Chitty on *Contracts* (24TH ed., 1977), Vol. I, paras. 961-1011.

[20] The "Chicago School" refers to a major tendency in neo-classical economics centred on past and present teachers at the University of Chicago: see Samuels, ed., *The Chicago*

*School of Poltical Economy* (1976), esp. at 5-7. In its application of economic analysis to legal phenomena its chief protagonist is undoubtedly Posner whose book, *Economic Analysis of Law* (2nd ed., 1977), and articles, "The economic approach to law", (1975) 53 *Texas L. Rev.* 757, and "Utilitarianism, economics and legal theory", (1969) 8 *Journal of Legal Studies* 109, provide perhaps the best introductions. The main vehicles for the work of the school in the legal field are the *Journal of Law and Economics* and the *Journal of Legal Studies*.

[21] See, e.g., Posner, *Economic Analysis of Law* (2nd ed., 1977).

[22] I.e., the principle of the legal enforceability of private bargains.

[23] See generally Kahn-Freund, in Ginsburg, ed., *Law and Public Opinion in England in the 20th Century* (1959), at 240-244.

[24] *Schroder Music Publishing Co.* v *Macaulay* (1974) 1 WLR 1308; (1974) 3 All ER 616; followed in *Clifford davis Management Ltd* v *WEA Records Ltd* (1975) 1 WLR 61; (1975) 1 All ER 237, but strongly criticized by Trebilcock, "The doctrine of inequality of bargaining power: post-Benthamite economcis in the Hose of Lords" (1976) 26 *U. Toronto L.J.* 359. The enactment of the Unfair Contract Terms Act 1977 may limit the development of this jurisprudence in that, in the important fields of exemption clauses, consumer contracts and standard form contracts, it removes the concepto of inequality of bargaining power from the field of common law development to that of statutory interpretation.

[25] *Supra*, Note 11.

[26] Figures from *The Government's Expenditure Plans 1981-82 to 1983-84* 1981; Cmnd. 8175) Table 1.2.

[27] *Ibid.*, Tables 1.7, 1.8.

[28] Some of the alternatives listed may in practice be excluded by moral or political considerations, or by specific legal obligations, but this is not important in relation to the present phase of the argument.

[29] This is, of course, not the only possible form of import quota system; alternatively, for example, a global quota could be allocated to importers on a first-served basis. While this might lead to some importers being able to obtain their whole requirement, and others none at all, the government woultd still need to police the activities of all importers to ensure that the quota was not exceeded.

[30] For an example see Breyer, "Analysing regulatory Failure: mismatches, less restrictive alternative, and reform", (1979) 92 *Harv. L. Rev.* 549, at 581: "The very fact that (taxes) do not prohibit an activity, or suppress a product *totally*, means that those with special needs and willingness to pay may obtain it. Taxes thus lessen the risk, present with standard setting, of working serious harm in an unknown special case".

[31] For some evidence in a United Kingdom context see Story, "An economic appraisal of the legal and administrative aspects of water pollution control in england and wales", in O'Riordan and D'Arce, eds., *Progress in resource Management and Environmental Planning*, (1979) Vol. 1, ch. 9.

[32] The phrase, but not the thought, is borrowed from Kadish and Kadish, *Discretion to Disobey: a Study of Lawful Departures from Legal Rules* (1973).

[33] I assume, in all uses for this example, that the government is indifferent to the country of origin of imported sets, a condition that may well not obtain in practice.

[34] One would expect a subsidy of £x on each domestic set to cause a greater increase in the purchase of domestic sets than would a tax £x on each foreign set, because there will be a number of consumers for whom the effect of the subsidy is to make them renounce the purchase of some quite different good in favour of a domestic set, and some for whom the effect of the tax is the purchase, not of a domestic set, but of some other good.

[34a] Cf. 2. Context *supra*.

[35] The fact that an objective is pursued by the use of higly specific and detailed measures

does not mean that it is precise. Standards in such areas as product quality and safety, workplace safety, etc., may be specified in exhaustive detail, but the objective may be no more precise than "accident reduction".

[36] Lord Lundy, who
"... from his earlies years
Was far too freely moved to tears"
rose to be in turn Secretary for India, the Colonies and War: but
"... if a Member rose to say
(as Members do from day to day)
Arising out of that reply...!
Lord Lundy would begin to cry".
This unstatesmanlike trait provoked a rapid slide from political favour, terminating in a posting as Governor of New South Wales. See Belloc, "Lord Lundy" in Roberts, ed., *Faber Book of Comic Verse* (2nd ed., 1974), at 315-317.

[37] Compare Parsons, *The System of Modern Societies* (1971), at 14 (influence, money and power as "generalized symbolic modes of social interschange").

[38] A helpful introduction to these arguments, illustrated from the cases, is to be found in Keir and Lawson, *Cases in Constitutional Law* (6th ed., 1980), at 69-111.

[39] Though the Bill of Rights did not purport to abolish all existing prerogative powers; some common-law powers to deploy force (e.g. in emergencies) remained: see *Att.-Gen* v *De Keyser's Royal Hotel* (1920) AC 508; *Burmah Oil Co. Ltd* v *Lord Advocate* (1965) AC 75.

[40] *Case of Proclamations* (1611) 12 Co. Rep. 74.

[41] For an example see Kenyon, *The Stuart Constitution* (1966), at 502-3.

[42] While the Court of Star Chamber was abolished in 1641, by the statute 16 Car. 1, cc. 10, 11, Many of its functions continued to be exercised in the Privy Council as such (or by its Cromwellian equivalent, the Council of State). The Scottish Privy Council survived until the Union of 1707.

[43] Parliamentary supremacy in taxation is specifically provided for in Bill of Rights, Art. 4, Claim of Right, Art. 7.

[44] Bill of Rights 1689, Arts. 1, 2.

[45] recent cases where this issue was raised but not resolved include R. v *Metropolitan Polcie Commissioner, ex parte Blackburn* (1968) 2 QB 108; R. v *Customs and Excise Commissioners, ex parte Cook* (1970) 1 WLR 450; *Inland revenue Commissioners* v *National Federation of Self-Employed and Small Businesses ltd* (1981) 2 All ER 93 (HL).

[46] *Supra*, Note 39.

[47] *Case of Mines* (1567) 1 Plowd. 310, at 315, 316.

[48] Petroleum (Prodution) Act 1934, s. 1.

[49] Continental Shelf Act 1964, s. 1(1).

[50] Cf. Firearms Act 1968.

[51] *Infra*, p. 246.

[52] Maitland, *Constitutional History of England* (1908), at 184.

[53] Specific appropriations were occasionally attached to tax statues, for example in 1624 (aids for the Protectorate) and 1665 (prosecution of the Dutch War), but Parliament greatly reduced its own capacity for control by granting both Charles II and James II large permanent revenues at the beginning of their reigns, which meant that they seldom needed to come back for additional sums.

[54] The Appropriation Act first appears in something resembling its modern shape in 1748, but strict annual appropriation is a later development. See generally Chester, *The English Administrative System 1780-1870* (1981), at 59-61, 185-191.

[55] (1611) 12 Co. Rep. 74.

[56] The statement of Lord Haldane in *Auckland Harbour Board* v *The King* (1924) AC 318, at 336, that it has been a principle of the British Constitution now for more than two centuries... that no money can be taken out of the Consolidated Fund... excepting under a distinct authorization from parliament itself...
ıpproaches nearest to being such a declaration, but is arguably *obiter*, being given in a new Zeland case, and in any event does not refer to *Legislative* authorization.

[57] See beer, *Treasury Control* (1956), at 51-52n.

[58] For details see Halsbury's *Laws of England* (4th ed.) Vol. 8, tit. "Constitutional law", para. 977.

[59] Elliott, however, has recently argued that "in certain circumstances legal action may complek the disbursement of money (voted under the Appropriation Act) if the withholding of money has the effect of frustating a statutory purpose": Appendix 18 to the Minutes of Evidence taken before the Select Committee on Procedure (Supply), para. 51 (1880-81) HCP 118-III.

[60] For some examples see, e.g., Industry Act 1972. Ss 7 and 8 authorize the granting of "financial assistance" to industry which "may be given on any terms and conditions, and by any description of investment or lending or guarantees, or by making grants" (s. 7(3)). Compare Part I and Sch. 1 of the Act, which provide only for the making of "grants" and specify, albeit in an incomplete way, some elements of the grantor-grantee relationship.

[61] Spending of these amounts may, of course, also be regulated by permanent spending legislation of the type referred to above.

[62] The example, and the argument, are developed in more detail in Daintith, "The functions of law in the field of short-term economic policy" (1976) 92 *LQR* 62, at 67-71.

[63] In classical legal usage, "dominium", meaning "ownership", may be contrasted with "imperium", meaning "supreme administrative power", as in "Omnia rex imperio possidet, singuli dominio": Seneca, *De Beneficiis* 7.5.1.

[64] See *Sutton's Hospital Case* (1612) 10 Co. Rep. 1a.

[65] In so far as the encouragement of this process is an end of government policy, this third type of law of direct relevance to our theme.

[66] *Supra*, p. 243.

[67] In Italy this offence carries a penalty of 6 months' to 2 years' imprisonment: Penal Code, Art. 341.

[68] See, in general, Hanbury and Maudsley, *Modern Equity* (10th ed., 1976), at 627-630; and for a recent example, *Re Brocklehurst, dec'd* (1978) Fam. 14.

[69] For a brief survey see VerLoren van Themaat, *Economic Law in the Member States in an Economic and Monetary Union: Interium Report* (1973, Commission of the European Communities Competition - Approximation of Legislation Series No 20), at 28-29.

[70] Industry Act 1975, s. 27 and Sch. 5.

[71] For general accounts of these events, see references *supra*, Note 4. Among users of the term "unconstitutional" were sir Geoffrey Howe, then Shadow Chancellor of the Exchequer, counsel to the john Lewis Partnership (a firm from which government contracts were withdrawn), and Professor Wade, for whom the governments's action was "a complete repudiation of primary constitutional principle": *op. cit. supra*, Note 4, at 56.

[72] Elliot *supra*, Note 59, seems to wish to lead us in this direction.

[73] "Law State and economy: The Industry Act 1975 in context" (1975) 2 Brit. Jl. Law and Society 103.

[74] Farjat, *Droit Economique* (2nd ed., 1982), *passim*.

[75] For a brief statement, with notes to sources, see Teubner, "Substantive and reflexive elements in modern law: toward the reconstruction of a theory of legal evolution" (1982) 16 *Law and Society Rev.* esp. at ?. Cf. Unger, *Law in Modern Society* (1976), at 192-223 ("The disintegration of the rule of law in post-liberal society").

# Retour à Costa
# La primauté du droit communautaire à la lumière du droit international

BRUNO DE WITTE
Florence

## I. Introduction: l'originalité de l'arrêt Costa

L'arrêt *Costa/ENEL*[1] de la Cour de justice européenne est souvent considéré un des piliers de l'ordre juridique communautaire: il aurait établi la règle de la primauté du droit communautaire sur le droit national. C'est là une affirmation un peu sommaire, qui ne réussit pas à justifier la célébrité pourtant méritée de cet arrêt. En effet, les tribunaux internationaux n'ont pas manqué, bien avant la naissance des Communautés, de proclamer et d'appliquer la primauté de la règle du droit des gens sur le droit national. Quand un tribunal international, au cours d'un litige qui lui est soumis, se trouve confronté avec deux règles contradictoires, une de droit international et l'autre de droit interne, il n'hésite pas à appliquer la première, et à lui donner ainsi la primauté. Parmi les nombreux exemples dans la jurisprudence internationale[2], on peut citer l'avis sur les *communautés gréco-bulgares*[3], proclamant la supériorité par rapport à la *loi* interne: "c'est un principe généralement reconnu du droit des gens que, dans les rapports entre puissances contractantes à un traité, les dispositions d'une loi interne ne sauraient prévaloir sur celles du traité"; ou encore l'avis sur le *traitement des nationaux polonais*[4], qui affirme cette primauté face aux *constitutions* nationales: "Un État ne saurait invoquer vis-à-vis d'un autre État sa propre Constitution pour se soustraire aux obligations que lui impose le droit international ou les traités en vigueur."

Cette primauté ne souffre pas d'exceptions. Elle n'existe pas seulement dans les rapports entre États, hypothèse classique que l'on retrouve dans les deux cas cités, mais s'applique à tout litige porté devant une juridiction internationale: quelle que soit la partie lésée – État, organisme international ou "simple" individu[5] –, jamais un État ne peut invoquer son propre droit face à la règle internationale.

Cette supériorité caractérise aussi *toutes* les sources du droit international; pour la règle conventionnelle, elle s'impose avec une évidence particu-

lière: en tant que règle commune aux contractants, elle ne peut se voir opposer des mesures unilatérales de la part d'une d'entre elles. Ce principe a été codifié par l'article 27 du traité de Vienne.

La Cour de justice européenne, de son côté, avait d'ailleurs suivi l'exemple des tribunaux internationaux, dans un arrêt obscur, antérieur à *Costa/ENEL*, l'arrêt *Humblet/État belge* de 1960[6]. Il s'agissait d'une procédure contentieuse, introduite par un fonctionnaire aux Communautés contre une mesure individuelle de l'administration fiscale belge, contraire au protocole d'immunité des fonctionnaires CECA. Dans sa décision, la Cour dit:

"qu'en effet, si la Cour constate dans un arrêt qu'un acte législatif ou administratif émanant des autorités d'un État membre est contraire au droit communautaire, cet État est obligé, en vertu de l'article 86 du traité CECA, aussi bien de rapporter l'acte dont il s'agit que de réparer les effets illicites qu'il a pu produire; que cette obligation résulte du traité et du protocole qui ont force de loi dans les États membres à la suite de leur ratification et qui l'emportent sur le droit interne"[7].

Rares sont les commentateurs qui ont noté ce passage de l'arrêt, et ceux-là ont des difficultés à en expliquer la "banalité", surtout en regard de l'imposante construction de l'arrêt *Costa* quelques années plus tard. Leur sentiment semble être que la Cour, dans *Humblet*, aurait en quelque sorte lâché une phrase sans trop y penser et n'aurait véritablement réalisé l'importance des enjeux que dans *Costa*[8]. En fait, l'affirmation de la primauté dans l'arrêt *Humblet* est très nette[9], malgré son caractère laconique. Puisque "la primauté est le fait de tout traité"[10], la Cour n'avait aucune raison de mettre en relief cette évidence, même si les traités communautaires ne la proclamaient pas explicitement.

Si alors l'arrêt Costa/ENEL est entré dans l'histoire de l'intégration européenne, c'est que, contrairement à Humblet, il a apporté une innovation par rapport aux canons classiques du droit international public. Dans cet arrêt, la Cour ne se borne pas à proclamer la primauté du droit communautaire, qui est chose admise, elle ajoute que ce droit, au plan national, "ne pourrait ... se voir opposer un texte interne quel qu'il soit"[11]. La Cour ne se prononce pas sur son *propre choix* entre deux normes (internationale et interne) dans un cas qui lui serait soumis; pour la première fois, elle se réfère au choix que doit faire le *juge national* dans une telle situation. La Cour impose à ce juge national l'obligation d'adopter la même position que lui, c'est-à-dire de reconnaître la primauté du droit communautaire. Ainsi se trouve élargie la *portée* du principe de primauté: à la simple *primauté internationale*, l'arrêt *Costa* ajoute ce qu'on pourrait appeler la primauté dans l'application interne ou *primauté interne*.

La distinction entre "primauté internationale" et "primauté interne" était connue de la doctrine internationaliste[12], bien que pas toujours dans ces mêmes termes. Mais seule la première était considérée une règle du

droit des gens; la seconde était laissée au droit public interne des États, qui pouvaient l'appliquer ou non, de manière discrétionnaire. Par contre, à part quelques auteurs qui ont exactement cerné l'apport innovateur de *Costa*[13], la distinction n'est guère explicitée dans la doctrine européiste, qui se contente généralement de dire que l'arrêt a consacré le principe de la "primauté" tout court. Cela peut mener à divers malentendus. Ainsi, certains s'étonneront de l'importance donnée à l'arrêt *Costa*, puisqu'il ne fait que répéter "une solution classique depuis longtemps consacrée par le droit international"[14]. D'autres, par contre, pèchent par l'excès inverse. Impressionnés par la nouveauté de la primauté interne, ils oublient l'existence de la primauté internationale dont elle dérive. Par là, la "primauté" devient un phénomène tout à fait particulier au droit communautaire, qui ne peut s'expliquer que par la nature différente de celui-ci par rapport au droit international désormais qualifié de "classique". La question de la primauté devient ainsi un des facteurs cruciaux du divorce doctrinal entre droit européen et droit international.

Dans le chapitre suivant, nous voudrions réagir contre cette évolution et démontrer, au contraire, que la nouveauté de *Costa* réside dans sa *forme* et non dans son *contenu*, et que la primauté interne, si elle est proclamée explicitement par la Cour de justice, existait déjà à l'état latent dans le droit international. La primauté d'application interne pourrait bien être, en fait, une conséquence nécessaire, mais pas toujours reconnue, de la primauté internationale d'un traité; c'est grâce au remarquable agencement juridictionnel du traité CEE, et au volontarisme de la Cour, que ce lien s'est enfin trouvé pleinement articulé.

## II. Le droit international et la primauté interne

En dehors des accords purement intergouvernementaux pas très nombreux[15], "le droit international a besoin pour son exécution du droit interne. Il ne peut se réaliser ni sans lui ni en dehors de lui"[16]. Afin de réaliser cette exécution interne, l'État reçoit, en vertu du principe de la primauté du droit international, l'obligation générale de conformer son ordre interne aux dispositions internationales par lesquelles il est lié[17]. Cependant, l'État dispose d'une certaine latitude dans le choix des voies et moyens pour s'acquitter de cette obligation de conformité. Dans le cadre de ce qu'on a qualifié de "structure décentralisée du droit international"[18], chaque État choisit ses propres règles d'exécution, mais celles-ci sont soumises à un conditionnement international: quand elles ne réussissent pas, dans un cas donné, et pour quelque motif que ce soit, à réaliser la conformité du droit interne à la norme internationale, il y a une violation du droit international qui peut être sanctionnée par la responsabilité internationale de l'État.

La marge de discrétion de l'État varie selon le type d'obligation. Le rapport du professeur Ago[19], qui a servi de base à la codification de la responsabilité internationale proposée par la Commission du droit international, distingue à cet égard entre *des obligations de comportement et des obligations de résultat*[20]. Les premières sont des obligations précises qui ne peuvent être exécutées de plusieurs façons. Aucun choix n'est laissé à l'État: il doit adopter (ou omettre) un comportement déterminé, sous peine de commettre un fait illicite et de voir naître immédiatement sa responsabilité internationale. Un exemple caractéristique est l'obligation d'adopter une convention portant loi uniforme.

Les obligations de résultat, en revanche, et elles sont les plus nombreuses, ne prescrivent pas de comportement précis. L'État peut arriver au résultat voulu par plusieurs voies. Il peut même rattraper une conduite initiale incompatible avec la règle internationale, par une autre action (du même ou d'un autre organe) plus conforme[21]. "De moyen en moyen, l'État dérape dans l'exécution de son obligation internationale, mais il ne choît réellement que quand tous les moyens disponibles ont été mis en œuvre, par l'intermédiaire de l'ensemble de ses organes, sans que soit atteint pourtant le résultat requis"[22]. Ainsi, l'obligation de résultat va toujours de pair, au plan procédural, avec la règle coutumière[23] de *l'épuisement des recours internes*. A l'origine, la règle est un arrangement de bon ordre, permettant d'éviter que les relations entre États – seuls compétents traditionnellement pour engager une action en responsabilité internationale – ne subissent le contrecoup de toute dispute entre un État et un individu étranger[24]. Plus généralement, l'action internationale doit rester un recours ultime, quand tous les moyens internes d'obtenir justice ont failli. Cette règle a également l'avantage de concevoir l'État comme une entité pluraliste, non monolithique, où un des organes est en mesure de réparer des violations éventuellement commises par un autre, où, en particulier, le pouvoir judiciaire peut corriger l'action de l'exécutif, et où les cours supérieures peuvent réformer les décisions des échelons inférieurs.

La marge de discrétion que représente pour l'État l'obligation de résultat n'existe pas seulement au niveau de nombreuses règles "concrètes" du droit international, mais également au niveau des règles "abstraites" relatives aux rapports de système entre droit international et droit interne. Ainsi, il est communément admis que les États sont libres de choisir entre les diverses méthodes de réception du droit international: *l'adoption*, automatique ou non, qui est liée à la doctrine moniste et introduit la règle internationale en tant que telle dans l'ordre interne; ou la *transformation*,
d'inspiration dualiste, qui recrée la règle internationale en une règle nationale avant qu'elle puisse être exécutée. Les deux méthodes sont considérées équivalentes au regard des exigences du droit international[25]; il suffit que son contenu soit valablement appliqué[26], dans quelque forme que ce soit.

La question de *l'effectivité* de la règle internationale dans l'ordre interne mène aussi à celle de son *rang*, qui nous occupe avant tout ici. La doctrine de l'adoption ouvre la porte à la reconnaissance de la supériorité du droit international dans l'ordre interne, mais ne l'implique pas nécessairement. Tant en théorie qu'en pratique, on trouve du monisme à primauté nationale comme du monisme à primauté internationale[27]. La doctrine de la transformation, elle, exclut la primauté: puisque la règle internationale a été transmuée en une règle interne, la question de son rang disparaîtra, à strictement parler. La norme transformée occupera, dans la hiérarchie des sources, le rang que possède l'acte dans lequel elle est coulée, ce qui sera, le plus souvent, une loi ordinaire. Les conflits de la loi d'origine internationale avec d'autres lois internes se résoudra alors selon la règle classique de succession dans le temps, *lex posterior derogat priori*. La primauté d'application de la norme internationale sera donc toujours accidentelle et précaire.

Par conséquent, une fois admis le libre choix entre transformation et adoption, il en suit logiquement que les États ont également la liberté de reconnaître ou non la primauté interne du droit international. Et en effet, une réponse négative à la question de la primauté ne mène pas nécessairement à un résultat incompatible avec la règle internationale. Le droit interne peut déjà être en conformité avec la règle; aussi longtemps qu'il n'existe effectivement, de loi nationale contraire au droit international, il n'y a pas de conflit possible, et la question de la primauté reste toute théorique. Même l'émanation d'une loi nationale qui contredit une disposition internationale, dans un pays appliquant la règle de "lex posterior", ne constituera pas encore violation du droit des gens. Cette loi peut, en effet, être à nouveau abrogée ou modifiée pour la conformer au droit international, avant qu'elle n'ait reçu d'application concrète. De même, le traité peut être modifié ou abrogé entre-temps. Enfin, la loi apparemment contraire peut être appliquée, dans la pratique administrative, de façon à sauvegarder le respect de la disposition conventionnelle.

Pour que naisse le fait illicite international, il faut donc que la potentialité de conflit se réalise, qu'une obligation internationale spécifique soit effectivement méconnue dans un cas concret d'application par les organes administratifs ou judiciaires de l'État. Rien ne s'oppose à ce que la responsabilité de l'État soit engagée par cette application. Selon l'article 6 de la codification de la responsabilité par la Commission du droit international,

> "le comportement d'un organe de l'État est considéré comme un fait de cet État d'après le droit international, que cet organe appartienne au pouvoir constituant, législatif, exécutif, judiciaire ou autre, que ses fonctions aient un caractère international ou interne et que sa position dans le cadre de l'organisation de l'État soit supérieure ou subordonnée".

Donc, les agissements d'organes exécutifs même subordonnés ou d'organes judiciaires même entièrement indépendants peuvent engager l'État.

Une thèse contraire confondrait les notions de "gouvernement" et d'"État".

Encore faut-il tenir compte de la règle d'épuisement des recours. L'obligation d'appliquer le droit international en cas de conflit est une obligation de résultat; l'État ne peut donc être tenu pour responsable que si le non-respect de la primauté constitue un fait illicite irréversible, c'est-à-dire lorsque sa juridiction de dernier ressort aura statué en ce sens.

Il faut pourtant immédiatement préciser que la règle de l'épuisement ne s'applique qu'aux *recours effectifs*, ceux dont une décision plus conforme au droit international peut raisonnablement être attendue. Cette réserve, généralement admise[28], se comprend aisément: la raison d'être de la règle de l'épuisement étant de permettre à l'État de rattraper le faux pas initial, la seule présence formelle d'un recours ne saurait suffire; il doit être capable de restaurer la conformité. De nombreux obstacles, de fait ou de droit[29], peuvent s'y opposer. Ainsi, l'applicant est exempté du recours si l'instance concernée ne juge pas dans un délai raisonnable ou n'offre pas de garanties d'impartialité. Il est également exempté lorsqu'il peut raisonnablement[30] estimer que l'organe de recours ne fera que répéter la décision précédente, soit parce qu'une *jurisprudence constante* exclut un résultat plus fidèle au droit international[31], soit parce que la première décision n'a fait qu'appliquer correctement une *règle impérative de son ordre interne*[32].

Or, à bien voir, la règle du "lex posterior", ou toute autre règle qui refuse à la norme internationale sa primauté, n'est que la variante "structurelle" d'une des deux dernières catégories. Elle sera "jurisprudence constante" lorsqu'elle trouve son origine dans une création jurisprudentielle, ce qui est, nous le verrons, le cas le plus fréquent. Normalement, la jurisprudence constante porte sur le *fond* de l'affaire; dans notre cas, elle porte sur la *question préliminaire* du rang du droit international, question qu'elle couvre de manière structurelle.

La règle de "lex posterior" sera au contraire "disposition impérative du droit interne" quand elle est exprimée dans un texte écrit, constitutionnel ou législatif. Puisqu'une telle règle vise précisément à couvrir chaque cas de concurrence entre une disposition nationale et internationale, elle sera nécessairement impérative. Une instance d'appel ne pourra donc que confirmer, ici aussi, la première décision rendue.

Nous pouvons donc conclure ainsi la question de la primauté interne selon le droit international: quand une *autorité nationale quelconque*, notamment et surtout *toute juridiction* quel que soit son degré, applique dans une affaire qui lui est soumise la règle nationale contraire au droit international, de manière consciente et en conformité avec la règle de conflit donnée par son droit interne[33], il n'y a plus d'espoir d'obtenir un redressement ultérieur au sein de cet ordre juridique et les recours internes doivent être considérés comme épuisés. Dès lors, il y a violation du droit international et la responsabilité de l'État est engagée[34]. Toute application de la

règle nationale contraire devient ainsi illicite et expose l'État à des sanctions internationales[35].

Au terme du raisonnement, il faut quelque peu revoir la prémisse selon laquelle l'État, en règle générale, est lié quant au résultat à atteindre, mais libre quant aux moyens d'y parvenir. En fait, et ce n'est pas surprenant, le but fixé déteint sur le choix des moyens et rend injustifiés certains d'entre eux. Ainsi, les États peuvent bien choisir dans l'abstrait la méthode de transformation, mais ne peuvent tirer de cette théorie ce qui est sa principale conséquence pratique: l'application de la règle "lex posterior" en cas de conflit. De même, le droit international n'impose pas la primauté interne en tant que règle abstraite, mais elle impose une *primauté d'application* au stade administratif ou judiciaire. Dans un système dualiste, le juge national se trouvera pris dans un dilemme: s'il applique la loi postérieure nationale, il expose son État à une sanction internationale; si, au contraire, il assure la primauté du droit international, il passe outre le mandat que lui donne son ordre interne et agit en organe de droit international, ce qui est une hérésie pour la doctrine dualiste.

Il reste également une barrière dualiste à franchir pour pouvoir affirmer que la formulation de l'arrêt *Costa/ENEL* correspond en tous points au contenu de la règle de primauté interne en droit international telle que nous venons de la présenter. En effet, la Cour de justice ne dit pas simplement que l'État italien violerait le traité lorsque le juge italien donnerait la primauté à la loi italienne postérieure. Elle dit que le droit communautaire "ne pourrait ... se voir opposer un texte interne quel qu'il soit", ou, en termes positifs, que le juge italien doit donner la primauté au droit communautaire. La question traditionnellement posée en termes de *responsabilité* est donc tournée en termes *d'obligation*[36]. C'est là une formulation inédite[37], mais aussi, pour un dualiste, un pas matériel que le droit international ne saurait franchir: dans le premier cas, le juge international prend simplement acte de la solution à laquelle est parvenu l'ordre interne et évalue celle-ci selon les prescrits du droit international; dans le deuxième cas, il s'immisce dans le processus juridictionnel interne et transgresse ainsi la barrière entre le système international et le système interne.

En fait, cela est une construction artificielle. Comme le montre bien K. Marek[38], lorsque le juge international statue en responsabilité, il n'évalue pas l'action de l'État de façon abstraite, mais examine toujours les actions concrètes des organes de l'État qui ont provoqué le fait illicite; il passe donc bien au crible le fonctionnement interne du système national. Que fait alors la Cour de justice européenne? Au lieu de juger les agissements internes après coup, comme le ferait toute autre juridiction internationale[39], elle impose une obligation pour le futur; dans les deux cas, la règle employée est la même: la primauté d'application du droit communautaire (ou international) sur le droit interne; dans le premier cas, elle est formulée

de manière *rétrospective*, sous forme de déclaration de responsabilité, dans le deuxième cas, de manière *prospective*, sous forme d'une obligation de faire. Wyatt exprime bien ce glissement[40]:

"Suppose 'the clock had been stopped' between the executive implementation of the discriminatory taxation, and judgment in the national court, and an international tribunal had been asked: what will the legal consequences be of a final court decision failing to give effect to this treaty provision? The tribunal would have answered: breach of the treaty. Now frame the question in terms of duty – what is the duty of the national courts? The answer would doubtless be: the duty of the courts is to give effect to the treaty."

## III. Les raisons du caractère innovateur de l'arrêt Costa

Si la déclaration d'une obligation équivaut à une déclaration de responsabilité, pourquoi les tribunaux internationaux n'ont-ils jamais fait la première? Pourquoi a-t-il fallu attendre la Cour de justice de Luxembourg et son arrêt *Costa* pour assister à cette innovation? On aura commencé à en soupçonner la raison. La Cour de justice a pu apporter cette réponse parce que la question lui a été posée dans les termes requis, grâce à l'agencement juridictionnel particulier du traité CEE.

Dans l'affaire *Costa/ENEL*, la Cour a été saisie par le Giudice Conciliatore de Milan sur base de l'article 177, qui prévoit notamment, comme on le sait, le renvoi préjudiciel en interprétation du droit communautaire par le juge national. C'est ce remarquable mécanisme juridictionnel qui a permis, selon l'expression de Wyatt[41], d'"arrêter la pendule" et de reformuler la doctrine traditionnelle de responsabilité. Grâce au renvoi, la Cour de justice peut intervenir au beau milieu de l'application interne du droit communautaire. Le juge national ne perd pas le pouvoir d'appliquer le droit, mais il doit le faire en tenant compte de l'interprétation authentique délivrée par la Cour de justice et des obligations que celle-ci peut comporter.

Cependant, la seule existence du recours préjudiciel en interprétation n'explique pas encore suffisamment l'innovation de *Costa*. Car déjà des juridictions internationales, fût-ce de manière exceptionnelle, avaient pu, dans des contextes spécifiques, se prononcer sur l'application en cours d'un traité dans l'ordre interne. Le cas le plus fameux est sans doute l'avis de la Cour permanente de justice internationale dans l'affaire de la *compétence des cours de Dantzig*[42]. Au cours de cette procédure, qui s'écarte nettement de l'action traditionnelle en responsabilité, la Cour avait à se prononcer sur la compétence des tribunaux de Dantzig à statuer sur les demandes basées sur le "Beamtenabkommen" entre la Pologne et Dantzig. Son intervention se situait donc forcément *avant* l'application interne du traité des fonctionnaires. Cependant, le point litigieux était uniquement de savoir si les cours de Dantzig pouvaient appliquer directement[43] ce traité, et non pas si, en cas de réponse positive, elles devaient reconnaître

l'autorité supérieure du traité face à une éventuelle loi interne contraire.

D'ailleurs, la Cour de justice aussi avait prononcé d'autres arrêts basés sur un renvoi préjudiciel, avant l'affaire *Costa*, sans formuler sa doctrine de primauté. Dans l'arrêt *Van Gend en Loos*[44], en particulier, il s'agissait d'un conflit potentiel entre la loi néerlandaise et l'article 13 du traité. Mais la primauté du droit communautaire n'était pas mise en question; le juge néerlandais de renvoi était prêt, sur base de l'autorisation expresse de l'article 66 de sa Constitution, à reconnaître cette primauté; la seule condition était que la disposition communautaire ait des effets directs, et c'est uniquement cette question-là qu'il posait à la Cour de justice. Celle-ci pouvait donc se dispenser d'articuler sa doctrine de primauté.

Dans *Costa*, la situation était différente. La question de la primauté y fut jetée sur le tapis par une objection préliminaire du gouvernement italien. Celui-ci contestait la recevabilité du renvoi préjudiciel par le Conciliatore de Milan: une interprétation des articles du traité CEE ne pouvait avoir pour ce juge la moindre utilité dans le cas présent, puisqu'il était de toute façon obligé d'appliquer la loi italienne de nationalisation du secteur électrique, que celle-ci soit contraire ou non au traité. Si la Cour de justice voulait établir sa compétence, elle devait donc nécessairement montrer que le renvoi avait un sens; et il ne pouvait avoir un sens que dans la mesure où le juge italien devait accorder la primauté au droit communautaire, nonobstant la règle de conflit "lex posterior" à laquelle son pays adhérait traditionnellement[45]. Bien sûr, la Cour aurait pu s'en tenir strictement à l'interprétation des articles du traité relatifs au cas qui lui était soumis par le juge milanais (les articles 102, 93, 53 et 37, paragraphe 2), mais elle aurait ainsi entériné une violation quasi certaine, dans un avenir proche, de la règle communautaire. Dans *Costa*, la Cour sort de la pure "interprétation" et s'occupe également de l'"application", mais c'est sans doute inévitable si l'on admet que la véritable fonction du recours préjudiciel en interprétation est de maintenir une certaine uniformité dans la mise en œuvre du droit communautaire[46]. Cette uniformité ne dépend pas seulement d'une interprétation unifiée des textes, mais autant des conditions d'application de ces textes.

L'originalité juridictionnelle du traité de Rome ne se limite d'ailleurs pas à l'article 177. Le recours en constatation des manquements des États est également une procédure "dépassant de loin les règles jusqu'à présent admises en droit international classique pour assurer l'exécution des obligations des États"[47]. Une des particularités de l'article 169, qui ne sera explicitement affirmée que dans un arrêt postérieur à *Costa*, est que cette procédure ne connaît pas la règle de l'épuisement des recours internes[48]. Une des justifications possibles de cette absence serait que la construction communautaire entraîne une obligation de solidarité particulière entre les partenaires et que la tolérance exprimée par la règle de l'épuisement doit céder la place à une conception plus exigeante qui sanctionne toute

infraction, même temporaire, même réparable, par un des organes de l'État, conception qui se trouve exprimée par l'article 5 du traité[49]. Cet argument téléologique nous mène droit au débat sur la nature spécifique des Communautés.

Une justification moins controversée pour l'absence de la règle de l'épuisement est donnée par la structure interne de la procédure de constatation de manquements. Au cours de la phase précontentieuse, la Commission, qui estime qu'un État a manqué à ses obligations communautaires, met celui-ci clairement devant son devoir et lui laisse d'amples possibilités de redresser la situation avant de mettre en marche la phase contentieuse devant la Cour[50]. La règle de l'épuisement perd donc sa raison d'être[51].

L'absence de la règle en droit communautaire y facilite – par rapport au droit international traditionnel – la démonstration de la règle de primauté interne. On ne doit plus prouver que la "lex posterior" (ou toute autre règle de conflit contraire à la primauté du droit international) constitue un "épuisement structurel" des recours internes.

## IV. Le fondement spécifiquement communautaire de la primauté dans l'arrêt Costa

Le raisonnement suivi jusqu'ici visait donc à démontrer que la nouveauté de l'arrêt *Costa/ENEL* est de nature surtout formelle et que la doctrine de la primauté interne se trouve déjà en germe dans le droit international. Or, quand on lit l'arrêt, on s'aperçoit que la Cour de justice n'a pas du tout suivi ce même chemin pour arriver à sa solution. Au contraire, elle ne se réfère au droit international public que pour mieux s'en démarquer, elle ne mentionne ni l'article 177 ni l'article 169 du traité et se fonde exclusivement sur une logique autonome, propre au droit communautaire, pour fonder la primauté interne de celui-ci, comme il apparaît de la conclusion du raisonnement:

"Le droit né du traité ne pourrait donc, *en raison de sa nature spécifique et originale*, se voir judiciairement opposer un texte interne quel qu'il soit."

Avant de rechercher, dans la section suivante, les motifs de ce choix des juges européens, il paraît utile de reprendre d'abord les divers arguments avancés pour démontrer la supériorité du droit communautaire. Ces arguments partiels préparent la conclusion globale, que nous venons de citer, de deux façons: ils tendent à démontrer à la fois la *primauté* du droit communautaire et sa *nature spécifique*, deux caractéristiques qui se retrouvent indissolublement liées dans la conclusion. D'où une certaine ambiguïté dans le raisonnement; le souci de prouver la nature propre de l'ordre

communautaire et de couper les ponts avec le droit international fait recourir la Cour à des arguments qui ne sont pas toujours pertinents au problème de la primauté. Citons d'abord, une fois de plus, les fameux paragraphes en question[52]:

1) attendu qu'à la différence des traités internationaux ordinaires, le traité CEE a institué un ordre juridique propre, intégré au système juridique des États membres lors de l'entrée en vigueur du traité et qui s'impose à leurs juridictions;

2) qu'en effet, en instituant une Communauté de durée illimitée, dotée d'institutions propres, de la personnalité, de la capacité juridique, d'une capacité de représentation internationale et plus particulièrement de pouvoirs réels issus d'une limitation de compétence ou d'un transfert d'attributions des États à la Communauté, ceux-ci ont limité, bien que dans des domaines restreints, leurs droits souverains et créé ainsi un corps de droit applicable à leurs ressortissants et à eux-mêmes;

3) attendu que cette intégration au droit de chaque pays membre de dispositions qui proviennent de source communautaire et, plus généralement, les termes et l'esprit du traité ont pour corollaire l'impossibilité pour les États de faire prévaloir, contre un ordre juridique accepté par eux sur une base de réciprocité, une mesure unilatérale ultérieure qui ne saurait ainsi lui être opposable;

4) que la force exécutive du droit communautaire ne saurait, en effet, varier d'un État à l'autre à la faveur des législations internes ultérieures, sans mettre en péril la réalisation des buts du traité visée à l'article 5, paragraaphe 2, ni provoquer une discrimination interdite par l'article 7;

5) que les obligations contractées dans le traité instituant la Communauté ne seraient pas inconditionnelles, mais seulement éventuelles, si elles pouvaient être mises en cause par les actes législatifs futurs des signataires;

6) que, lorsque le droit d'agir unilatéralement est reconnu aux États, c'est en vertu d'une clause spéciale précise (articles 15, 93, paragraphe 3, 223 à 225 par exemple);

7) que, d'autre part, les demandes de dérogation des États sont soumises à des procédures d'autorisation (articles 8, paragraphe 4, 17, paragraphe 4, 25, 26, 73, 93, paragraphe 2, troisième alinéa, et 226 par exemple) qui seraient sans objet s'ils avaient la possibilité de se soustraire à leurs obligations au moyen d'une simple loi;

8) attendu que la prééminence du droit communautaire est confirmée par l'article 189 aux termes duquel les règlements ont valeur "obligatoire" et sont "directement applicables dans tout État membre";

9) que cette disposition, qui n'est assortie d'aucune réserve, serait sans portée si un État pouvait unilatéralement en annihiler les effets par un acte législatif opposable aux textes communautaires;

10) attendu qu'il résulte de l'ensemble de ces éléments qu'issu d'une source autonome, le droit né du traité ne pourrait donc, en raison de sa

nature spécifique originale, se voir judiciairement opposer un texte interne quel qu'il soit, sans perdre son caractère communautaire et sans que soit mise en cause la base juridique de la Communauté elle-même;

11) que le transfert opéré par les États, de leur ordre juridique interne au profit de l'ordre juridique communautaire, des droits et obligations correspondant aux dispositions du traité entraîne donc une limitation définitive de leurs droits souverains contre laquelle ne saurait prévaloir un acte unilatéral ultérieur incompatible avec la notion de Communauté".

**a) Le transfert d'attributions**

Le premier argument pour fonder la primauté se trouve annoncé dans l'attendu 2, qui analyse la construction communautaire et souligne en particulier que les États ont "limité leurs droits souverains" par un "transfert d'attributions" à la Communauté, qui a doté celle-ci de "pouvoirs réels". Ce paragraphe vise avant tout à démontrer l'originalité du système communautaire que la Cour vient d'affirmer dans l'attendu 1. En même temps, il veut sans doute indiquer que la question de la primauté se pose en toute son acuité dans le cas de la Communauté, parce que celle-ci crée, par l'exercice de ses pouvoirs, un droit secondaire ou dérivé destiné à s'appliquer au niveau interne.

Dans l'attendu 3 alors, la Cour semble tirer argument de cet exercice de pouvoirs propres pour fonder la primauté:

"que cette intégration au droit de chaque pays membre de dispositions qui proviennent de source communautaire" a "pour corollaire l'impossibilité pour les États de faire prévaloir ... une mesure unilatérale ultérieure".

Cet argument a été répété, plus clairement encore, dans un autre arrêt de la Cour[53]. Il a également remporté un certain succès auprès de la doctrine[54]. Pourtant, il semble particulièrement déconcertant à première vue. En effet, le transfert de compétences opéré par un traité n'est opposable aux États signataires que dans la mesure où le traité lui-même est doté de cette primauté; le droit dérivé qui naît par l'exercice de ces compétences n'a pas de légitimité propre, de primauté autonome, mais participe de la primauté de l'acte conventionnel, et seulement dans la mesure où il reste dans le cadre tracé par cet acte[55]. Or, il s'agit précisément ici[56] de démontrer la primauté du *traité* CEE; fonder celle-ci sur l'attribution de pouvoirs que le traité comporte, c'est mettre la charrue avant les bœufs.

Mais il s'est développé également une interprétation plus ambitieuse de cet argument. Selon cette version, l'attribution de pouvoirs opérée dans le cadre des traités communautaires a entraîné pour les États "une limitation définitive de leurs droits souverains"[57]. Ce démembrement de la souverai-

neté des États a permis à la Communauté de constituer sa propre sphère de souveraineté, à l'intérieur de laquelle les États ont perdu tout pouvoir de légiférer. Cela a "pour corollaire l'impossibilité pour les États de faire prévaloir... une mesure unilatérale ultérieure" (*Costa*, attendu 3). Selon cette théorie, le mot d'"impossibilité", employé par la Cour, n'est pas synonyme d'"illicéité", mais doit être pris dans son sens fort; une loi postérieure contraire au droit communautaire n'a que l'apparence d'une loi; elle est, en fait, nulle *ab initio*. Puisqu'il n'y a plus de véritable conflit, on ne peut plus parler de primauté, mais plutôt d'un partage de souveraineté, tant la Communauté que l'État régnant sans partage dans son propre domaine.

Comme la plupart des thèses basées sur la notion de "souveraineté", celle-ci est fortement théorique. La pratique, elle, montre que de telles dispositions nationales contraires ne sont pas du tout considérées comme nulles, ni par les autorités nationales, qui les appliquent sans hésitations, ni même par la Cour de justice. Celle-ci peut déclarer illicite l'application d'une disposition nationale, mais elle ne peut l'abroger[58]. Elle n'aurait d'ailleurs pas le pouvoir de mettre à exécution une telle annulation.

Enfin, on voit bien l'intérêt que peut avoir cette thèse pour assurer la prééminence du droit dérivé; celle-ci ne devra plus s'appuyer sur la primauté du traité, mais aura sa propre légitimité. En revanche, on ne voit toujours pas comment cette thèse peut expliquer la primauté du traité lui-même.

### b) L'applicabilité directe

Dans l'attendu 8 de l'arrêt Costa, la Cour se réfère à l'article 189 aux termes duquel les règlements sont "directement applicables" et en déduit, dans l'attendu suivant, qu'ils doivent obtenir la prééminence sur le droit national.

Le terme d'"applicabilité directe" est un des plus riches du droit européen, mais également un de ceux qui ont le plus prêté à équivoque. Il peut référer à deux concepts très différents. Applicabilité directe peut vouloir dire *réception automatique* du droit communautaire, en tant que droit communautaire, dans l'ordre interne, sans nécessité de transformation préalable; dans un deuxième sens, elle signifie la capacité d'une norme communautaire d'attribuer directement un droit individuel et/ou d'être *applicable en justice* sans mesures d'exécution interne préalables. La doctrine internationaliste distingue nettement entre la première notion, qui est une règle *structurelle* relative au *mode d'introduction* du droit international dans l'ordre interne, et la seconde, qui indique le caractère "self-executing" d'une norme *individuelle, une fois introduite* en droit interne[59]. Certains auteurs[60] ont voulu distinguer, en droit communautaire égale-

ment, entre *l'effet direct* (ou caractère self-executing) qui caractériserait un grand nombre d'articles du traité et du droit dérivé, et l'*applicabilité directe* proprement dite (ou réception automatique) qui serait le fait des seuls règlements, en vertu de l'article 189.

La jurisprudence conduit à relativiser un peu cette distinction[61], mais elle reste utile pour notre propos, puisque les deux concepts ont tour à tour été utilisés pour fonder la primauté interne.

Aucun des deux arguments ne réussit cependant à convaincre. Quant à l'*appplicabilité directe* du règlement, l'argument employé par la Cour, nous avons vu que la question de l'effectivité du droit international doit être distinguée de celle de son rang[62]. La réponse à l'une ne préjuge pas toujours l'autre. Certains pays reconnaissent d'ailleurs l'applicabilité directe de certaines décisions internationales[63], sans leur accorder la primauté sur la loi postérieure interne. D'ailleurs, même si l'applicabilité directe impliquait la primauté, cela ne vaudrait que pour les règlements, et non pas pour les autres sources de droit communautaire, et notamment pour les traités qui ne sont pas directement applicables dans le sens où le terme est employé ici[64].

Quant à l'effet direct, il a sans doute un lien étroit avec la primauté, mais alors en sens inverse. Le caractère self-executing d'une norme internationale est la condition pour qu'elle soit appliquée, et donc aussi pour qu'elle soit appliquée de préférence à une norme nationale[65]. Le contraire n'est pas vrai: l'effet direct n'implique pas nécessairement la primauté. Une norme peut être parfaitement appliquée par le juge ou administration, sans mesures nationales intermédiaires, et néanmoins se heurter occasionnellement à une norme interne contraire, qui reçoit la priorité en vertu de la règle "lex posterior" ou autre.

### c) Le caractère commun et l'uniformité

Le troisième point d'appui pour fonder la primauté porte en un terrain plus familier pour l'analyse internationaliste. Dans l'attendu 3, la Cour dit que "les termes et l'esprit du traité ont pour corollaire l'impossibilité pour les États de faire prévaloir, *contre un ordre juridique accepté par eux sur une base de réciprocité, une mesure unilatérale* ultérieure...". Les mots soulignés n'expriment rien d'autre que le principe de "*pacta sunt servanda*", le caractère obligatoire du traité pour ses signataires, qui implique sa primauté sur le droit interne[66]. Ce thème est précisé plus en détail dans les paragraphes 5 à 7, qui précisent que les États membres, en dehors des autorisations expressément accordées par le traité lui-même, ne peuvent prendre de mesures unilatérales.

L'argument se retrouve sous la plume de nombreux spécialistes du droit communautaire, mais rares sont ceux qui le relient à la règle "pacta sunt

servanda" du droit international; on préfère généralement parler du "caractère commun" du droit communautaire, suggérant par ces mots qu'il s'agit là à nouveau d'une spécificité de ce droit[67]. Cette nouvelle terminologie n'empêche pas que le fond de l'argument reste très classique et ne permet pas automatiquement d'arriver à la conclusion, beaucoup moins classique, que le droit communautaire doit prévaloir dans le for interne. Afin de justifier cette portée élargie de la primauté, la Cour ajoute un considérant qui est peut-être le véritable cœur de l'arrêt, l'attendu 4:

"que la force exécutive du droit communautaire ne saurait, en effet, varier d'un État à l'autre à la faveur des législations internes ultérieures, sans mettre en péril la réalisation des buts du traité visée à l'article 5, paragraphe 2, ni provoquer une discrimination interdite par l'article 7".

Ce qu'on peut appeler l'argument d'*uniformité* n'a pas d'existence autonome. Il s'appuie sur la démonstration du "caractère commun" (la primauté internationale), mais en même temps l'approfondit en montrant que ce caractère commun s'accompagne nécessairement d'une uniformité dans l'application interne (la primauté interne). Pour fonder l'exigence d'uniformité, la Cour, à nouveau, évite tout recours au droit des gens, mais avance deux articles du traité: les articles 7 et 5, paragraphe 2.

*L'article 7* interdit "toute discrimination exercée en raison de la nationalité". L'application d'une mesure nationale contraire à la règle communautaire aboutira bien à une *disparité de situations*: tandis que la règle communautaire est bloquée par la mesure nationale dans l'État A, elle continue de s'appliquer pleinement dans l'État B. Mais elle ne provoquera que rarement une *discrimination*, car on accepte généralement de caractériser celle-ci comme un "traitement différencié imputable à *une* personne ou autorité". Dans notre cas, l'administration ou le juge national appliqueront la règle nationale contraire tant aux citoyens de l'État qu'aux autres ressortissants communautaires; ce n'est que lorsque cette règle contient elle-même une distinction injustifiée entre ces deux catégories que l'on pourra parler de discrimination au sens de l'article 7. Sinon, on déplace abusivement les limites de ce concept: chaque illégalité territorialement limitée devrait être analysée comme une discrimination[68] [69].

Second argument de texte, l'*article 5, paragraphe 2*, du traité CEE selon lequel les États membres "... s'abstiennent de toutes mesures susceptibles de mettre en péril la réalisation des buts du présent traité". Cet article peut être vu comme une simple reformulation de l'obligation prise par les parties contractantes d'un traité d'exécuter celui-ci de bonne foi, principe classique du droit international. Mais il peut être considéré, alternativement, comme formulant une véritable obligation de "loyauté communautaire" qui se rapproche de la loyauté fédérale exigée des membres d'un État fédéral. La Cour de justice a estimé, dans une affaire postérieure, qu'un État, en omettant d'abroger une législation nationale contraire au

droit communautaire, même si elle n'était *pas effectivement appliquée*, manquait à ses obligations communautaires définies dans l'article 5[70]. A fortiori pourrait-on affirmer que l'application d'une loi nationale contraire constitue violation.

Ces deux arguments de texte ont été très peu utilisés par la suite, tant par la Cour que par la doctrine. Celle-ci s'est efforcée de chercher une base plus solide pour l'exigence d'uniformité dans l'*esprit* plutôt que dans les *termes* du traité. Certains auteurs influents ont ainsi élevé l'uniformité en une véritable "exigence existentielle" de la construction communautaire[71]. Une telle dramatisation n'est pas corroborée par la réalité qui montre que le droit communautaire ne s'applique pas partout dans la même mesure, pour diverses raisons, et notamment parce que la primauté interne n'est pas reconnue dans certains États ou parce que le droit communautaire est mal interprété. Néanmoins, la Communauté, à ce qu'il semble, survit toujours. Il faudrait donc nuancer et dire que l'*existence* n'est menacée qu'à partir d'un *certain degré* de diversité dans son application. Loin d'être un principe existentiel immuable, l'uniformité est un concept relatif, dépendant des circonstances. La Cour elle-même reconnaît d'ailleurs aux États membres une marge importante de diversité en ce qui concerne les règles procédurales nationales permettant la revendication des droits communautaires[72]. La non-uniformité, tolérée, de ces règles mène, en pratique, à une non-uniformité du droit matériel qui est sans doute plus grave que celle qui résulte du non-respect de la primauté.

## V. Les motifs de la théorie de la nature spécifique

L'analyse de l'arrêt *Costa* et des auteurs qui ont développé son inspiration montre donc deux phénomènes intéressants: d'une part, la Cour et la doctrine dominante n'inscrivent pas leur démarche, visant à fonder la primauté interne, dans la continuité avec le droit international. Même si certains arguments (le "caractère commun" et l'"uniformité") se réfèrent implicitement au droit international, ce lien n'est pas mis en valeur. D'autre part, ils avancent des arguments supplémentaires (le transfert d'attributions et l'applicabilité directe) dont le lien avec le problème de la primauté est fortement spéculatif.

Ces deux phénomènes s'intègrent, selon nous, dans le cadre d'un objectif unique de la Cour: *démarquer le droit communautaire du droit international afin de mieux assurer sa primauté*. Et la raison pour laquelle la Cour poursuit cet objectif pourrait bien être cherchée dans l'*autonomie institutionnelle* des États membres et la façon dont celle-ci s'exprimait concrètement en 1964. La Communauté n'est pas (encore?) un État fédéral qui peut faire exécuter ses règles de droit, au besoin par la contrainte. Au contraire, tout comme le droit international, le droit communautaire laisse

aux États un rôle essentiel dans la mise en œuvre de ses règles. L'application effective de ce droit et la sanction de son respect passent par les organes de l'État. La Cour de justice peut bien préciser les obligations qui leur incombent en vertu du traité, et c'est ce qu'elle fait dans le cas présent en formulant l'exigence de primauté interne. Mais ce sont en définitive les États qui, seuls, décident d'exécuter ces obligations, sans autre sanction éventuelle que la condamnation pour manquement aux articles 169 à 171 CEE. Pour perfectionné qu'il soit par ailleurs, le système juridictionnel européen ne réussit pas là à dépasser le cadre classique du droit international.

Cela étant, la Cour de justice devra se soucier non seulement de "dire le droit", mais aussi de la réception que ses arrêts auront au niveau national; elle devra, pour compenser la faiblesse de ses moyens de contrainte, développer une véritable "politique judiciaire" en direction des États membres. Puisqu'elle ne peut *forcer* les autorités nationales à accepter son interprétation, pourtant authentique, du traité, elle devra essayer de les *persuader*. Cela vaudra tout spécialement pour une question aussi essentielle que la primauté interne, qui est un aspect particulièrement chatouilleux des sensibilités nationales[73]. La mise en œuvre du principe de primauté sera donc nécessairement bidimensionnelle[74]. "Aussi longtemps qu'il n'existe pas de structure fédérale, une solution uniforme du conflit entre les ordres juridiques communautaire et nationaux ne peut être trouvée que si chacun de ces ordres juridiques contient une règle de conflit équivalente"[75].

Replaçons-nous maintenant dans le contexte concret de l'arrêt *Costa/ENEL* en 1964. Nous avons vu que le gouvernement italien contestait la recevabilité du renvoi préjudiciel par le Giudice Conciliatore de Milan, puisque celui-ci devait en tout état de cause – même contre une disposition du traité – appliquer la loi italienne[76]. Le gouvernement pouvait s'appuyer directement sur l'arrêt que la cour constitutionnelle italienne venait de rendre entre les mêmes parties, *Costa/ENEL*, et sur renvoi de la même juridiction inférieure[77]. La cour constitutionnelle y avait répété la doctrine dualiste bien ancrée en Italie: le droit national est la seule source de légalité dans l'ordre interne; toute règle d'origine internationale ne peut y opérer qu'après sa transformation en une norme italienne, et c'est la nature de cet acte de transformation qui détermine sa place dans la hiérarchie des normes[78]. La cour italienne n'avait vu aucune raison d'accorder au traité de Rome un traitement privilégié par rapport à d'autres conventions internationales; le traité CEE ayant été transformé par une simple loi, rien n'empêchait une loi postérieure – *in casu* la loi de nationalisation du secteur électrique du 6 septembre 1962 – d'y déroger.

Le coup de semonce de l'arrêt *Costa* italien n'était que la manifestation concrète d'une menace plus diffuse. En Allemagne d'abord – autre pays d'ancienne tradition dualiste –, le même danger pointait, même s'il n'y

avait pas encore été concrétisé. Dans la doctrine en tout cas dominait l'opinion selon laquelle le droit communautaire, en tant que branche du droit international conventionnel, ne pouvait comme lui avoir de validité interne qu'après transformation, avec comme conséquence logique l'application de la règle "lex posterior" en cas de conflit[79].

En France et en Belgique également, la primauté interne du droit européen était loin d'être assurée en 1964, mais pour d'autres raisons. Ces pays adoptaient une position plutôt moniste et la Constitution française reconnaissait même explicitement la primauté du droit international dans son article 55. Les difficultés y provenaient plutôt de la doctrine de séparation des pouvoirs, en vertu de laquelle les juges ne se sentaient pas autorisés à censurer l'œuvre du législateur national, pour quelque motif que ce soit, et même si c'était pour pouvoir appliquer la règle internationale. La supériorité théoriquement reconnue à cette dernière, en France du moins, restait lettre morte face à une loi postérieure nationale.

Deux puissantes doctrines, le *dualisme* et la *séparation des pouvoirs*, concouraient donc pour faire craindre une attitude fermée des juges nationaux devant l'impératif communautaire[80]. Seuls les Pays-Bas et le Luxembourg assuraient la primauté au droit international, et donc vraisemblablement au droit communautaire, le premier pays par l'article 66 de sa Constitution, déjà mentionné[81], et le second par une jurisprudence orientée en ce sens[82].

Il est permis de penser que la Cour de justice, en rendant son jugement dans l'affaire *Costa*, ait été sensible au danger de morcellement et de désintégration que ce traitement généralement insatisfaisant des normes de droit international infligeait à la construction communautaire. Si besoin en était, l'avocat général Lagrange l'avait encore clairement rappelé dans ses conclusions sous l'arrêt. A la fois consciente des limites imposées par l'autonomie institutionnelle des pays membres et soucieuse d'atteindre le résultat désiré (la primauté interne du droit communautaire), la Cour a pu être tentée de manier à la fois le bâton et la carotte. D'une part, elle proclame fermement la priorité de la règle communautaire dans l'ordre interne. D'autre part, plutôt que d'attaquer de front les doctrines vénérables et solidement enracinées que sont le dualisme et la séparation des pouvoirs et de montrer leur inadéquation à l'évolution juridique, elle a sans doute voulu contourner l'obstacle.

L'avocat général avait, dans ses conclusions, posé une alternative radicale: "Si le juge national trouve que la primauté ne peut être appliquée dans le cadre des normes constitutionnelles de son pays, il n'y a que deux solutions: modifier la Constitution ou dénoncer le traité", sauf, ajoutait-il cependant, si on pouvait "découvrir le moyen constitutionnel"[83], malgré tout, pour assurer la pleine application du traité. C'est précisément à cela que s'emploie la Cour. Si on pouvait laisser entendre au juge national que le droit communautaire se différencie, de par sa nature, du droit interna-

tional public, peut-être la question du *rang* de ce droit dans l'ordre interne recevrait-elle aussi une réponse *sui generis*, plus favorable. Peu importe alors si les États membres maintiennent leur doctrine dualiste ou séparationniste quant à la relation entre droit international et droit interne, du moment que le droit communautaire peut échapper à ce sort grâce à sa nature spécifique.

Selon nous, il y aurait donc en quelque sorte une double logique dans l'arrêt *Costa*: une logique apparente, qui déduit la primauté du droit communautaire de sa nature propre, et une logique intérieure, exactement inverse, qui déduit la nature spécifique du droit communautaire de la nécessité de lui assurer la priorité au plan interne. Cette seconde logique a été pleinement articulée ailleurs que dans l'arrêt. Un excellent exemple est l'intervention du président de la Commission, M. Hallstein, quelque temps après:

"On a tenté çà et là d'assimiler le droit communautaire au droit international public général ... Si cette opinion était vraie, le juge allemand et le juge italien devraient en effet – comme ils doivent le faire à l'égard du droit international public – appliquer toujours le droit national, si flagrante que soit son opposition avec le droit communautaire. En revanche, le juge néerlandais – pour ne citer que l'autre extrême – devrait toujours, d'après l'article 67 de la Constitution néerlandaise, opérer inversement, à savoir laisser inappliquée la loi nationale qui va à l'encontre d'une règle de droit européen. Cela est inacceptable ... La Communauté n'a pas d'infrastructure administrative, pas de pouvoir direct de coercition, pas d'armée et pas de police. Son unique instrument, sa seule arme, c'est le droit qu'elle fixe. Il est clair que sa mission serait au plus haut point menacée, et en définitive mise en échec, si cet unique moyen de mettre en œuvre les objectifs communautaires perdait son caractère obligatoire uniforme dans tous les États membres"[84].

Précisons tout de suite que la Cour "laisse entendre", dans l'arrêt *Costa*, que l'ordre communautaire a rompu les amarres avec le droit international public, mais ne l'affirme pas explicitement. En faisant cela, elle serait d'ailleurs entrée en contradiction flagrante avec sa propre décision *Van Gend en Loos* de l'année précédente, où elle avait décrit la Communauté comme un "ordre juridique nouveau de droit international"[85]. Pour la doctrine "spécifiste", ce n'était là qu'une simple "remarque quelque peu malencontreuse", sans suite[86]. Selon les mots d'un autre auteur[87]:

"Si un jour la Cour de justice a cru voir dans le droit communautaire une certaine forme de droit international, elle s'est hâtée de souligner qu'il s'agissait d'un ordre de droit des gens assorti de caractéristiques très particulières; ensuite, la Cour a corrigé son erreur possible en déclarant, sans laisser de doute, que le droit communautaire constitue un ordonnancement juridique propre, particulier et interne des Communautés."

L'arrêt *Costa* ne laisse-t-il vraiment aucun doute? L'attendu 1 reste en tout cas fort prudent: "attendu qu'à la différence des traités internationaux ordinaires, le traité CEE a institué un ordre juridique propre". En affirmant que le traité CEE n'est pas un traité international *ordinaire*, la

Cour ne sous-entend-elle pas qu'il continue d'appartenir à la catégorie des traités internationaux? En fait, cet attendu 1 ne fait, semble-t-il, qu'amorcer le virage qui, de la formulation de *Van Gend en Loos*, mène à la formulation de l'attendu 10 de *Costa*, où la Cour conclut que "le droit du traité ... en vertu de sa *nature spécifique originale...*".

En résumé, sans exclure en toutes lettres l'appartenance internationale des Communautés, la Cour n'en laisse pas moins entendre qu'une telle interprétation pourrait bien être donnée à sa pensée. C'est la doctrine qui s'est chargée de pleinement articuler la doctrine de la spécificité du droit communautaire[88].

## VI. La Communauté: ordre juridique "sui generis" ou sous-système du droit international?

Sans reprendre ici tout le débat sur la nature juridique de la Communauté, on voudrait simplement noter qu'il a été parfois faussé par une vision beaucoup trop étriquée du droit international public. Celui-ci a beaucoup évolué, et notamment depuis l'apparition du phénomène de l'organisation internationale. Le droit des gens ne constitue pas un tout monolithique aux règles uniformes. On y peut au contraire distinguer entre un *droit relationnel* classique et un *droit institutionnel* aux caractéristiques fort différentes[89]. Ce droit institutionnel est constitué par un ensemble d'organisations internationales régies par un certain nombre de règles qui leur sont propres et qui constituent autant d'"ordres juridiques particuliers".

Cette dernière notion n'a rien de choquant ou de révolutionnaire en droit des gens. En fait, il faut "considérer l'ordre international général comme un système universel qui inclut la multiplicité des ordres juridiques particuliers comme sous-systèmes"[99]. Les règles propres de ces sous-systèmes que sont les organisations internationales divergent parfois sensiblement des solutions classiques du droit international[91], notamment quant à leurs structures et leurs compétences. Ainsi, ces organisations possèdent-elles des organes propres dotés d'une certaine autonomie d'action et qui ne représentent pas nécessairement (tous) les États membres. Les fonctions de ces organisations s'exercent souvent selon un mode qui préserve au maximum la libre disposition des États membres, mais parfois aussi elles peuvent prendre la forme d'un véritable *pouvoir réglementaire*. Ces "règlements" ou "décisions" peuvent se définir par trois qualités cumulatives: il s'agit d'abord d'actes *externes*, destinés aux États et non au fonctionnement interne de l'organisation; ce sont des mesures *unilatérales*, prises par les organes de l'institution sans acceptation préalable de la part des États[92]; elles sont enfin *obligatoires*, contrairement aux recommandations et autres exemples de "soft law" international[93].

La théorie de la spécificité radicale de la construction communautaire paraît sous-estimer la capacité de développement du droit international. Plusieurs des indices proposés dans *Costa*, et repris par la doctrine, pour justifier le caractère sui generis de la Communauté (ainsi que sa primauté) ne sont en fait pas tellement spécifiques.

"En instituant une Communauté de durée illimitée, dotée d'institutions propres, de la personnalité, de la capacité juridique, d'une capacité de représentation internationale et plus particulièrement de pouvoirs réels issus d'une limitation de compétence ou d'un transfert d'attributions des États à la Communauté"[94], les signataires des traités européens n'ont pas innové. Sans doute, le *transfert des fonctions* à d'autres organisations internationales n'atteint pas la même ampleur; on peut affirmer également que, dans le cas des autres organisations, ces compétences sont d'ordre essentiellement technique, alors qu'elles sont véritablement politiques dans le cas des Communautés. Mais dans les deux cas, c'est bien la méthode fonctionnaliste qui a inspiré les fondateurs à confier à des organes autonomes certaines compétences bien délimitées. L'objectif visé par cette méthode fonctionnaliste peut être différent: simple coopération dans un cas et *intégration* dans l'autre. Mais peut-on attacher des conséquences juridiques à ces arrière-pensées idéologiques?

Deuxième argument souvent avancé, l'*effet direct* n'est pas non plus l'apanage exclusif du droit européen. Le phénomène est traditionnellement présenté dans la littérature internationale sous un autre nom ("caractère self-executing"), mais il est fondamentalement identique dans les deux cas[95]. Nier que le droit des gens contient des dispositions à effet direct est une position dogmatique qui relègue ce droit dans le domaine purement intergouvernemental[96]. En ordre subsidiaire, il est parfois affirmé que le droit communautaire accomplit un renversement de la présomption. D'exception en droit international, l'effet direct serait devenu la règle en droit communautaire[97]. En fait, l'effet direct attire seulement plus l'attention en droit européen, parce qu'il y est reconnu par la Cour de justice de manière centralisée à travers la procédure de l'article 177, alors que, pour le reste du droit international, son appréciation est laissée aux tribunaux internes et est donc plus dispersée[98]. En cette matière comme en d'autres, c'est par son agencement juridictionnel plus que par son contenu matériel que se distingue le droit communautaire.

Quant à l'*applicabilité directe* au sens étroit où cette notion a été définie plus haut[99], elle n'était pas inconnue avant la naissance des Communautés. Seulement, le choix de reconnaître une efficacité immédiate à la règle nationale, sans aucun acte de réception interne, était normalement laissé à la discrétion des États. Certains ordres juridiques reconnaissaient une telle applicabilité directe à des règlements internationaux[100]. Le fait nouveau, ou relativement nouveau[101], est que, à travers l'article 189 CEE, *tous* les États membres reconnaissent l'applicabilité directe du règlement commu-

nautaire, indépendamment du caractère moniste ou dualiste de leur ordre juridique[102]. Voilà pourquoi l'argument de l'applicabilité directe impressionne surtout dans les pays de tradition dualiste, qui ne connaissaient pas du tout ce phénomène avant. Ce "Durchgriff" ("pénétration")[103], cette "immédiateté"[104] sont pour eux les particularités décisives du système communautaire.

Les remarques de la Cour de justice dans *Van Gend en Loos* sont très intéressantes:

"La Communauté constitue un nouvel ordre juridique de droit international, au profit duquel les États ont limité, bien que dans des domaines restreints, leurs droits souverains et dont les sujets sont non seulement les États membres, mais également leurs ressortissants."

On ne saurait mieux dire que ni l'attribution de pouvoirs ni l'effet direct ne permettent de fonder le caractère spécifique, non international, de l'ordre juridique communautaire.

Le système communautaire offre d'autres points originaux que ceux suggérés par la lecture de l'arrêt *Costa*. Soerensen[105] dit fort justement que sa particularité principale par rapport aux règles connues du droit international est sans doute son système juridictionnel[106]. Mais la véritable question est celle-ci: est-ce que le droit communautaire, malgré ses nombreux traits spécifiques, garde sa place, en tant qu'"ordre juridique particulier", au sein du droit international; ou est-ce que sa spécificité est telle qu'elle lui fasse quitter la sphère du droit international et qu'on doive parler désormais d'un ordre juridique "sui generis", ni international, ni étatique? Pour répondre à cette question, il ne suffit pas de rejeter, comme nous l'avons fait, tous les indices partiels de spécificité. Même si tous les éléments séparés ne sont pas décisifs, leur accumulation et leur importance *quantitative* pourraient permettre au droit communautaire d'accomplir un saut *qualitatif* et de sortir de l'orbite du droit international[107].

Il arrive que des ordres juridiques étatiques trouvent leur origine historique dans des accords internationaux, qui se transforment en constitutions internes et échappent ainsi à l'ordre juridique international[108]. C'est une telle métamorphose que nombre d'auteurs croient déceler dans le cas des Communautés:

"Après leur entrée en vigueur, les traités instituant les Communautés européennes doivent être considérés comme des actes constitutifs des Communautés possédant en quelque sorte un caractère constituant. Comme tels, ils cessent d'appartenir au droit international classique"[109].

Dans la même veine, Ipsen[110] décrit le traité comme un acte d'intégration qui n'a fait qu'emprunter la forme du traité pour constituer les Communautés.

Une telle analyse peut relever du domaine du souhaitable, mais ne correspond guère à la réalité communautaire. On peut désigner la Commu-

nauté, afin de résumer en une formule tous ses traits distinctifs, comme une *organisation supranationale*, mais il faut garder présente la relativité de ce terme. Un auteur comme Schermers[111] propose de distinguer entre organisations intergouvernementales et organisations supranationales, mais ajoute aussitôt qu'il s'agit là de deux types idéaux, entre lesquels s'échelonnent les organisations concrètement existantes. La Communauté européenne se situe sans doute au bout du continuum, mais présente néanmoins toujours un savant dosage d'éléments d'intégration communautaire et d'éléments de coopération interétatique. En particulier, comme le montre l'analyse de Weiler[112], les progrès du degré de supranationalité au plan *normatif* – et la primauté interne en est un des aspects principaux – se sont accompagnés d'un recul correspondant au plan *décisionnel*.

Le processus décisionnel au sein de la Communauté montre que la volonté des États ne s'est pas éteinte par la signature des traités, mais continue de dominer l'évolution de l'ordre juridique communautaire. Surtout, les États membres ont gardé le contrôle de la question, cruciale pour notre propos, de l'amendement des traités; la procédure de modification des traités communautaires ne diffère guère d'autres organisations internationales. Sans doute les institutions véritablement communautaires (Parlement et Commission) y sont-elles associées, mais la décision finale est du ressort des États membres et de l'organe communautaire composé de leurs représentants, le Conseil. Bien sûr, à côté de la révision formelle, l'article 235 CEE s'est développé en une véritable forme de révision larvée. Mais là encore, cette évolution n'atteste pas la prédominance d'une volonté communautaire propre; elle ne s'est faite que par l'assentiment, et même sous l'impulsion des États, et peut être interrompue quand ils le désirent[113].

On peut conclure que l'ensemble des États membres reste plus fort que la Communauté et que la condition fondamentale pour pouvoir classer les Communautés européennes parmi les organisations internationales – la persistance de la volonté des États fondateurs – est donc bien toujours présente[114].

## VII. Les suites de "Costa"

Au-delà de la *validité* de la théorie de la nature spécifique du droit communautaire, que nous venons d'examiner, il convient de s'interroger sur son *utilité*. Est-ce que son objectif présumé – persuader les États membres d'attribuer la primauté au droit communautaire sans qu'il faille revoir leur attitude négative envers le droit international – a été atteint dans les années qui ont suivi l'arrêt *Costa*? Nous ferons pour cela un rapide survol de l'évolution des jurisprudences nationales, qui montre globalement des progrès considérables dans la reconnaissance de la primauté interne, mais en empruntant des chemins divers[115].

En *Allemagne* et en *Italie*, comme le note J. V. Louis, "la thèse de la spécificité prévaut à cause même de la conception traditionnellement dualiste des rapports entre droit international et droit interne. Si l'on voulait reconnaître un statut privilégié au droit communautaire, il fallait apporter des justifications propres à la nature de celui-ci"[116]. Dans les pays auxquels la théorie de la nature propre était surtout destinée, les cours ont donc saisi la perche tendue par la Cour de Luxembourg et ont ainsi pu accommoder le maintien de leur dualisme envers le droit international, avec une ouverture aux nécessités communautaires.

En dernière analyse cependant, cette solution spécifique pour le droit européen trouve sa légitimité, tant en Italie qu'en Allemagne, dans une disposition constitutionnelle interne, et non pas, comme l'aurait voulu la Cour de justice, uniquement dans le système communautaire lui-même. Tant l'article 24 de la loi fondamentale allemande que l'article 11 de la Constitution italienne permettent de transférer des pouvoirs étatiques à des organisations supranationales. Logiquement, ces articles se réfèrent seulement au moment de la conclusion et de la ratification d'accords créant des organisations internationales, et non pas au problème postérieur de l'effet et du rang de ces accords ou du droit secondaire créé par ces organisations[117]. Pourtant, tant la cour constitutionnelle allemande[118] qu'italienne[119] ont vu dans ces articles des "leviers" permettant – pour le seul cas de la Communauté européenne – de transpercer la cuirasse dualiste. Dans cette brèche, le droit communautaire peut librement s'engouffrer. Puisque ce droit émane d'une institution à laquelle a été transférée une portion de la souveraineté nationale, le juge national ne peut le soumettre à aucun contrôle et doit lui assurer la primauté[120].

Plusieurs difficultés demeurent cependant dans cette approche. Premièrement, si la théorie du transfert de souveraineté peut bien justifier l'"intangibilité" des actes pris par l'organisation en cause, elle a du mal à assurer le même statut à l'acte constitutif, le traité lui-même. Dans un de ses arrêts, la cour constitutionnelle allemande a trouvé une solution astucieuse, mais qui paraît un peu artificielle: parmi les actes secondaires intangibles pris en vertu du traité figurent aussi les arrêts de la Cour de justice. Si maintenant celle-ci oblige les juges nationaux à reconnaître la primauté interne du traité, ceux-ci doivent obéir à cette interprétation obligatoire[121].

La deuxième difficulté de cette construction est le fait que sa légitimité repose sur une disposition constitutionnelle interne. Les articles 24 et 11, respectivement, ont bien "ouvert la porte" au droit communautaire, mais celui-ci reste limité par les dimensions de cette porte. Le transfert de souveraineté autorisé par la Constitution ne comporte pas le pouvoir de porter atteinte aux principes de base de cette Constitution, et notamment aux droits fondamentaux. En conséquence, tant le *Bundesverfassungsgericht* que la *Corte Costituzionale* se sont réservé le pouvoir de contrôler la

conformité des actes communautaires aux règles constitutionnelles essentielles[122]. Il faut y ajouter également, en Italie seulement cette fois-ci, la doctrine *Simmenthal,* en vertu de laquelle la cour constitutionnelle, tout en reconnaissant la primauté, entend se réserver – à l'exclusion des autres juridictions italiennes – le droit de l'appliquer face à la loi postérieure. Selon la Cour, puisque la primauté du droit communautaire repose sur le transfert de souveraineté autorisé par l'article 11, elle fait désormais partie du "bloc de constitutionnalité", le respect duquel doit être assuré par la seule cour constitutionnelle[123].

En *France* et en *Belgique*, l'évolution a été très différente. Les juridictions belges et les juridictions ordinaires[124] en France ont, depuis 1964, sauté le pas et oublient désormais leur déférence à la loi nationale dans l'hypothèse où elle contraste avec une norme communautaire ou toute autre norme de droit international. S'il n'est pas douteux que le développement communautaire a constitué le facteur déterminant dans ce revirement jurisprudentiel, il est frappant de constater que les Cours de cassation de ces deux pays étendent le bénéfice de la primauté à l'*ensemble* du droit international, sans favoriser le droit européen. En France, c'est sur la base de l'article 55 de la Constitution – enfin pleinement appliqué – que la Cour de cassation a élaboré sa nouvelle doctrine dans l'affaire *Cafés Jacques Vabre*[125]. La Cour belge, au contraire, n'avait pas, dans son revirement, de point d'ancrage constitutionnel, puisque la Constitution belge est muette sur les rapports entre droit international et droit interne. La voie était libre ainsi pour le remarquable arrêt *Le Ski*[126], où la Cour a fondé la primauté du droit international sur la nature de ce dernier, et non sur une quelconque disposition interne. Assurer la primauté à l'ensemble des règles internationales n'était pourtant pas évident. Le cas d'espèce relevait du droit communautaire; la Cour aurait donc pu suivre la Cour de justice dans l'arrêt *Costa* et baser la primauté de la règle communautaire sur sa spécificité, et ainsi sauver sa jurisprudence traditionnelle par rapport au reste du droit international.

Parmi les autres pays membres, la *Grèce*, dans son article 28,1, de la Constitution, reconnaît la primauté de l'ensemble des traités internationaux, sans privilégier les traités communautaires. Au Royaume-Uni, en *Irlande* et au *Danemark*, la théorie de la spécificité a joué un rôle secondaire. Ces pays, tous trois dualistes envers le droit international, ont pris des mesures législatives afin de permettre l'application directe et effective du droit communautaire, mais sans parvenir – jusqu'à nouvel ordre – à assurer à celui-ci la primauté face à une loi postérieure contraire.

En résumé, si l'évolution de la reconnaissance de la primauté a été généralement favorable, la fortune de la théorie spécifiste a été plutôt mitigée. Certaines juridictions se sont converties à la primauté sans recours à elle. D'autres, notamment en Italie et en Allemagne, ont accueilli la doctrine, sans pourtant aboutir à une primauté claire et nette du droit

communautaire. D'où un certain effet paradoxal: en amenant les juridictions nationales à trouver le "moyen constitutionnel", à exploiter toutes leurs ressources constitutionnelles internes, la Cour ne les a pas suffisamment convaincues que la primauté du droit communautaire, avant d'être une concession étatique, est d'abord une exigence communautaire (et internationale). En outre, on ne peut sous-estimer les conséquences négatives qu'a pu avoir cette théorie pour le développement du droit international. Celui-ci sert trop souvent de repoussoir pour mieux assurer la pleine application du droit communautaire, alors qu'il aurait pu, au contraire, bénéficier de l'effet d'entraînement causé par cette branche dynamique.

## VIII. Conclusion

La règle bien connue du droit international selon laquelle les États sont obligés quant au résultat à atteindre, mais libres quant au choix des moyens, est souvent interprétée de façon un peu sommaire; on ne peut ranger, sous l'étiquette de "moyens", tous les aspects de l'efficacité interne des traités. Les États peuvent sans doute choisir entre les méthodes de transformation ou d'adoption du droit international; ils peuvent même déduire de là le rang théorique de la règle d'origine internationale dans l'ordre interne. Mais ils ne peuvent *appliquer* leur droit interne en cas de conflit avec le droit des gens. Ce dernier exige donc, sous peine de responsabilité internationale, une primauté d'application interne.

La solution apportée par l'arrêt *Costa/ENEL* en matière de primauté du droit communautaire s'inscrit donc parfaitement dans la foulée du droit international. Pourtant, cet arrêt suggère une coupure entre ces deux systèmes juridiques. Cette doctrine de la spécificité s'explique sans doute par le souci d'éviter l'application au jeune droit communautaire des canons traditionnels défavorables au droit international dans la plupart des pays. Il n'en reste pas moins que s'opère ainsi une confusion théorique, qui n'est pas sans conséquences néfastes pour le développement du droit international, tout en n'arrivant pas à vaincre toutes les résistances contre la primauté du droit communautaire.

"On ne sert pas une branche de droit en discréditant l'arbre dont elle est issue"[127]. Il est temps, dans l'intérêt des deux systèmes, de rapprocher le droit communautaire du droit international public. On s'apercevra alors que le droit européen "n'est, en quelque sorte, que le 'fer de lance' d'une nouvelle conception de l'État et, partant, d'une nouvelle conception du rôle du juge. Dans la mouvance du droit communautaire, nous assistons actuellement à une prise de conscience par le juge de ce qu'on pourrait appeler l''au-delà' de l'État, c'est-à-dire à une prise de conscience par le juge du conditionnement international de l'existence et de l'action des collectivités politiques nationales"[128].

## Notes

[1] Affaire 6-64, Recueil XI, p. 1149.

[2] Voir la liste donnée par L. Cavare, *Le droit international public positif*, Paris, 1967, T.I., p. 178 et suiv.

[3] CPJI, série B, n° 17, p. 32.

[4] CPJI, série A/B, n° 44, p. 24.

[5] Ce dernier cas se présente par exemple devant la Commission et la Cour européenne des droits de l'homme, lors de l'application individuelle de l'article 25 de la convention.

[6] Affaire 6-60, Recueil VI, p. 1125.

[7] Affaire 6-60, citée, p. 1146.

[8] Voir, par exemple, G. Bebr, *Development of Judicial Control of the European Communities*, The Hague-Boston-London, 1981, p. 635, qui parle d'un début "pas exactement très prometteur".

[9] H. P. Ipsen, *Europäisches Gemeinschaftsrecht*, Tübingen, 1972, p. 296, y voit à tort une primauté conditionnelle. Les mots "force de loi", employés par la Cour, sont sans doute malencontreux; mais dans le contexte, ils ne signifient pas une caractérisation hiérarchique ("force de loi = de même rang qu'une loi formelle"), mais plutôt l'efficacité interne de la norme ("force de loi = droit effectivement applicable dans l'ordre interne").

[10] M. Lagrange, "La primauté du droit communautaire sur le droit national", in *Droit communautaire et droit national*, Semaine de Bruges 1965, Bruges, 1965, p. 23.

[11] Affaire 6-64, citée, p. 1160.

[12] Voir déjà, par exemple, F. Morgenstern, "Judicial Practice and the Supremacy of International Law", in *British Year Book of International Law*, 1950, p. 42-92.

[13] Par exemple E. Stein, "Toward supremacy of Treaty Constitution by Judicial Fiat: On the Margin of the Costa Case", in *Rivista di Diritto Internazionale*, 1965, p. 24: "... it is perhaps the first time in history that a court established by an international treaty has asserted its power to determine with effect not only in the 'international' (or Community) legal order but also in national law, the hierarchical value of the very norm to which it owes its existence".

[14] D. Carreau, "Droit communautaire et droits nationaux: concurrence ou primauté? La contribution de l'arrêt Simmenthal", in *Revue trimestrielle de droit européen*, 1978, p. 393.

[15] Par exemple, une obligation conventionelle de négocier.

[16] K. Marek, "Les rapports entre le droit international et le droit interne à la lumière de la jurisprudence de la Cour internationale de justice", *Revue générale de droit international public*, 1962, p. 263.

[17] Pour tous, I. Brownlie, *Principles of Public International Law*, 2nd ed., Oxford, 1979, p. 36.

[18] A. Verdross & B. Stimma, *Universelles Völkerrecht*, Berlin, 1976, p. 68.

[19] Annuaire de la Commission de droit international, 1977, vol. II (première partie), p. 3-48.

[20] Le droit civil connaît une distinction homonyme, mais au contenu fort différent. En droit civil, l'obligation de résultat est la plus exigeante; en droit international, c'est l'inverse. Voir à ce sujet J. Combacau, "Obligations de résultat et obligations de comportement. Quelques questions et pas de réponse", in *Mélanges offerts à Paul Reuter. Le droit international: unité et diversité*, Paris, 1981, p. 193 et suiv. Cet auteur met aussi, fort justement, l'accent sur les difficultés à appliquer en pratique cette distinction théorique; voir *infra*, p. 263.

[21] Rapport Ago, cité, paragraphe 41.

[22] J. Combacau, *op. cit.*, p. 188.

[23] Affaire *Interhandel*, CIJ, Recueil 1959, p. 27: "La règle selon laquelle les recours internes doivent être épuisés avant qu'une procédure internationale puisse être engagée est une règle bien établie du droit international coutumier".

[24] La règle de l'épuisement ne vaut pas, par conséquent, dans les obligations interétatiques (E. Jimenez De Arechaga, "International Responsability", in M. Soerensen (ed.), *Manual of Public International Law*, London-New York, 1968, p. 545-546).

[25] Voir, par exemple, I. Seidl-Hohenveldern, "Transformation or Adoption of International Law into Municipal Law", in *International and Comparative Law Quarterly*, 1963, p. 88-124.

[26] La liberté de choix de États peut cependant être limitée par des dispositions conventionnelles expresses. Pour nous en tenir au cas qui nous occupe, le traité CEE contient une *obligation de transformation* pour une catégorie de normes secondaires, les directives, et une *interdiction de transformation* pour une autre catégorie, les règlements.

[27] Voir notamment P. Guggenheim, *Traité de droit international public*, I, Genève, 1967, p. 54 et suiv.

[28] D. R. Mummery, "The Content of the Duty to Exhaust Local Judicial Remedies", in *American Journal of International Law*, 1964, p. 395-404. H. Rolin, "Le contrôle international des juridictions internes", in *Revue belge de droit international*, 1967, p. 20. B. Vitanyl, "International Responsibility of States for their Administration of Justice", in *Netherlands International Law Review*, 1975, p. 163.

[29] Pour cette distinction, voir E. Jimenez De Arechaga, *op. cit.*, p. 588.

[30] D. R. Mummery, *op. cit.*, p. 401, essaie de préciser mieux ce critère de rationalité: l'applicant serait exempté lorsque le coût impliqué par un recours supplémentaire excède la possibilité d'un résultat satisfaisant.

[31] Affaire du *Chemin de fer Panevezys-Saldutiskis*, CPIJ, série A/B, n° 76, p. 18; affaire du *S.S. Lisman*, Reports of International Arbitral Awards 1939, p. 1773. Voir également la jurisprudence de la Commission européenne des droits de l'homme, par exemple affaire 27/55, *X République fédérale d'Allemagne*, Annuaire de la convention européenne des droits de l'homme, I, 1956, p. 138, et affaire 514/59, *X Autriche*, Annuaire de la convention européenne de droits de l'homme, III, 1960, p. 196. Les applications de *Lawless* et de *Handyside* ont été déclarées recevables par la Cour européenne des droits de l'homme, sans qu'il y ait eu appel au niveau national. L'absence de réception de la convention dans les États concernés (Irlande et Royaume-Uni respectivement) rendait ineffectif tout recours supplémentaire basé sur un article de la convention.

[32] H. Urbanek, "Das völkerrechtverletznede nationale Urteil", in *Österreichische Zeitschrift für öffentliches Recht und Völkerrecht*, 1958-1959, p. 232: "wenn es aufgrund der in dem Verletzerstaat anzuwendenden Gesetze von vornherein feststeht, daß die Rechtsmittelinstanzen an dem völkerrechtlichen Unrecht nichts ändern können". Voir aussi le *rapport Ago*, cité, paragraphe 42.

[33] Une autorité nationale peut également méconnaître la véritable règle de conflit du for, en appliquant la règle du lex posterior, alors qu'une disposition écrite ou une jurisprudence constante attribuent la primauté au droit international. Une telle violation "accidentelle" du droit des gens n'entraînera pas l'épuisement des recours internes, car elle pourra être corrigée en appel.

[34] A cet égard, on peut noter qu'une *faute* dans le sens traditionnel n'est pas requise pour la constitution de la responsabilité internationale; cette dernière peut fort bien être causée par un défaut de la structure institutionnelle ou du système judiciaire de l'État, sans qu'il y ait aucune intention subjective de nuire (E. Jimenez De Arechaga, *op. cit.*, p. 536). Contra, B. Vitanyi, *op. cit.*, p. 157, qui estime qu'il ne peut y avoir responsabilité pour l'administration de la justice qu'en cas d'intention délibérée de porter préjudice.

[35] Cela n'empêche pas, bien sûr, la règle nationale de continuer à sortir ses effets dans l'ordre interne. Le juge international peut déclarer une *illicéité*, mais ne peut frapper de *nullité*. F. Morgenstern, *op. cit.*, p. 91: "Municipal law in violation of international law may have

legal effects within the restricted sphere of the state concerned, but its enforcement is an illegality from the point of view of international tribunals."

[36] D. Wyatt, "New Legal Order, or Old?", in *European Law Review*, 1982, p. 133 et 145.

[37] P. Guggenheim, *op. cit.*, p. 66: "La Cour permanente de justice internationale, lorsqu'elle s'est trouvée devant une telle situation, n'a toutefois jamais condamné expressément l'application abusive du droit national. Elle s'est toujours bornée à constater en pareil cas que l'application du droit national avait eu pour conséquence la violation d'une règle du droit des gens."

[38] K. Marek, *op. cit.*, p. 267 et suiv.

[39] Et comme le ferait la Cour de justice elle-même dans le cadre du recours en manquement des articles 169 à 171 du traité CEE.

[40] D. Wyatt, *op. cit.*, p. 133.

[41] Voir note 40.

[42] Recueil CPJI, série B, n° 15.

[43] Est c'est pour sa définition de l'applicabilité directe que cet avis est devenu célèbre.

[44] Affaire 26-62, Recueil IX, p. 1.

[45] Voilà aussi pourquoi l'apport essentiel de l'arrêt, l'élabora-doctrinale de la primauté interne, se retrouve curieusement dans la question préliminaire, et non dans le fond de l'arrêt. Certains en ont conclu que le principe de primauté n'était qu'un "obiter dictum"!

[46] Dans l'arrêt *Bosch*, la première décision rendue sur base de l'article 177, la Cour affirme que le renvoi préjudiciel a pour objet de réaliser "une harmonisation des jurisprudences" (affaire 13-61, Recueil VIII, p. 89, attendu 12).

[47] Affaire 20-69, *Gouvernement de la République italienne Haute Autorité*, Recueil VI, p. 692. Cet arrêt se réfère à l'article 88 du traité CECA, mais vaut tout autant pour l'article 169 CEE.

[48] Affaire 31-69, *Commission Italie*, Recueil XVI, p. 33: "L'existence de voies de droit ouvertes auprès des juridictions nationales ne saurait préjudicier, à aucun égard, à l'exercice du recours visé à l'article 169, les deux actions poursuivant des buts et ayant des effets différents."

[49] Sur cet article, voir encore *infra*, p. 271. Il suffit de mentionner ici que certains auteurs déduisent de cet article exactement le contraire, c'est-à-dire qu'il laisserait les États membres entièrement libres de choisir la procédure de mise en œuvre de leurs obligations (G. Pau, "Sui limiti di rilevanza del diritto comunitario nel sistema giuridico italiano", in *Rivista di Diritto Internazionale*, 1978, p. 278.

[50] Sur cette phase précontentieuse, voir, avec des références instructives à la pratique, Ph. Cahier, "Les articles 169 et 171 du traité instituant la Communauté économique européenne à travers la pratique de la Commission et la jurisprudence de la Cour", in *Cahiers de droit européen*, 1974, p. 5-15.

[51] Voir *supra*, p. 260.

[52] Affaire *Costa ENEL*, citée; dans le recueil des arrêts, ces attendus ne sont pas numérotés. Nous avons ajouté ces numéros pour faciliter les références dans la suite de l'exposé.

[53] Affaire 48-71, *Commission République italienne*, Recueil 1972, p. 535: "que l'attribution, opérée par les États membres à la Communauté des droits et pouvoirs correspondant aux dispositions du traité, entraîne, en effet, une limitation définitive de leurs droits souverains, contre laquelle ne saurait prévaloir l'invocation de dispositions de droit interne de quelque nature qu'elles soient".

[54] C. Sasse, "The Common Market: Between International and Municipal Law", in *Yale Law Journal*, 1965-1966, p. 742 et suiv.; G. Bebr, *op. cit.*, p. 646.

[55] (annulé).

[56] Tant dans l'affaire *Costa* que dans l'affaire 48-71, il s'agissait d'un conflit entre une loi italienne et un article du traité, et non pas du droit dérivé.

[57] Affaire 48-71, citée.

[58] (annulé).

[59] Voir notamment P. De Visscher, "Les tendances internationales des constitutions modernes", in *Recueil des Cours*, vol. 80, 1952; M. Waelbroeck, *Traités internationaux et juridictions internes dans les pays du Marché commun*, Paris, 1969; A. Koller, *Die unmittelbare Anwendbarkeit völkerrechtlicher Verträge und des EWG-Vertrags im innerstaatlichen Bereich*, Bern, 1971, p. 58-68.

[60] J. Winter, "Direct Applicability and Direct Effect: Two Distinct and Different Concepts in Community Law", in *Common Market Law Review*, 1972, p. 425; J. P. Warner, "The Relationship between European Community Law and the National Laws of Member States", in *Law Quarterly Review*, 1977, p. 349; H. Schermers, *Judicial Protection in the European Communities*, Deventer, 1976, paragraphe 175.

[61] Voir à ce sujet R. Kovar, "L'intégrité de l'effet direct du droit communautaire selon la jurisprudence de la Cour de justice de la Comunauté", in *Das Europa der Zweiten Generation - Gedächtnisschrift für Christoph Sasse*, Kehl-Strasbourg, 1981, I, p. 151-169, et J. Steiner, "Direct Applicability in EEC Law - a Chameleon Concept", in *Law Quarterly Review*, 1982, p. 229-248.

[62] Voir *supra*, p. 261. L'argument de l'applicabilité directe est notamment rejeté, dans la doctrine européiste, par C. Sasse, *op. cit.*, p. 739, et L. J. Constantinesco, *Das Rechts der Europäischen Gemeinschaften*, Baden Baden, 1977, I, p. 669-670.

[63] Voir *infra*, p. 277.

[64] Dans chacun des États membres, le traité a été reçu dans l'ordre interne par un acte formel.

[65] En niant à la règle internationale son caractère self-executing, les autorités nationales trouvent parfois un moyen commode pour éviter de se prononcer sur sa primauté. Voir M. G. Marcoff, "Les règles d'application indirecte en droit international", RGDIP, 1976, p. 413.

[66] Voir *supra*, p. 257.

[67] Voir notamment F. Dumon, "L'afflux européen dans les droits et les institutions des États membres des Communautés européennes", *Cahiers de droit européen*, 1965, p. 19; J. Mertens De Wilmars, "La jurisprudence de la Cour de justice comme instrument de l'intégration communautaire", *Cahiers de droit européen*, 1976, p. 140; P. Pescatore, "Droit communautaire et droit national selon la jurisprudence de la Cour de justice des Communautés européennes", *Dalloz* 1969, Chr. 182-183; C. Constantinides-Megret, *Le droit de la Communauté économique européenne et l'ordre juridique des États membres*, Paris, 1967, p. 99.

[68] On peut faire le parallèle avec le droit public interne: lorsqu'une autorité administrative locale maintient, en face de la loi nationale, sa propre réglementation contraire, on ne parlera pas de discrimination mais d'illégalité ou excès de pouvoir.

[69] On a parfois élargi l'argument: plutôt que l'article 7 spécifiquement, c'est le principe général d'*égalité* devant la loi qui exige l'uniformité (J. Mertens De Wilmars, *op. cit.*, p. 140); mais cela n'est plus un argument; cette égalité n'est rien d'autre que l'uniformité qu'il s'agit de justifier...

[70] Affaire 167-73, *Commission République française*, Recueil 1974, p. 359.

[71] P. Pescatore, *op. cit.*, p. 183: "L'existence même de la Communauté est mise en cause dès lors que l'ordre juridique communautaire ne peut pas se réaliser avec des effets identiques et avec une efficacité uniforme sur l'ensemble de l'aire géographique de la Communauté." Pour H. Kutscher, "Community law stands and falls by its uniform validity and application in all the Member States."

[72] Voyez notamment les arrêts 33-76 (*Rewe Zentralfinanz Landwirtschaftskammer für das Saarland*) et 61-79 (Denkavit), et les analyses de G. Bebr, *Remedies for Breach of Communi-*

*ty Law*, rapport au 9e Congrès FIDE, Londres, 1980, et J. H. H. Weiler, *Supranational law and the supranational system: Legal structure and political process in the European Community*, Institut universitaire européen, Florence (dactylographié), 1982, p. 479-500.

[73] R. Kovar, "L'effectivité interne du droit communautaire", in *La Communauté et ses États membres*, Liège-La Haye, 1973, p. 222: "L'effectivité du droit communautaire dépend ainsi largement de la qualité des solutions juridiques dégagées dans l'ordre étatique. Ce fait essentiel a surtout été perçu à propos de la sanction judiciaire de la supériorité du droit des Communautés."

[74] J. H. H. Weiler, *op. cit.*, p. 74.

[75] J. A. Frowein, "Europäisches Gemeinschaftsrecht und Bundesverfassungsgericht", in *Bundesverfassungsgericht und Grundgesetz*, 1976, II, p. 197.

[76] Voir *supra*, p. 265.

[77] Cour constitutionnelle, arrêt n, 14/64 du 24 février 1964.

[78] Pour cette conséquence de la doctrine dualiste, voir *supra*, p. 261.

[79] Pour cette position orthodoxe, voir notamment H. J. Schlochauer, "Das Verhältnis des Rechts de Europäischen Wirtschaftsgemeinschaft zu den nationalen Rechtsordnungen der Mitgliedstaaten", in *Archiv des Völkerrechts*, 1963, p. 1.

[80] P. Pescatore, "Rôle et chance du droit et des juges dans la construction de l'Europe", in *Revue internationale de droit comparé*, 1974, p. 16-18.

[81] Voir *supra*, p. 265. Selon cet article 66, "les dispositions législatives ne sont pas applicables si leur application n'est pas conciliable avec des dispositions d'accords liant un chacun, qu'ils soient antérieurs ou postérieurs".

[82] Conclusions sous l'affaire *Costa*, citée, p. 1180.

[83] *Ibidem*, p. 1181.

[84] Intervention publiée dans *Revue trimestrielle de droit européen*, 1965, p. 250.

[85] Affaire 26-62, citée.

[86] G. Bebr, *Development of Judicial Control of the European Communites, op. cit.*, p. 636.

[87] J. L. Iglesias Buigues, "La nature juridique du droit communautaire", *Cahiers de droit européen*, 1968, p. 527.

[88] Outre les auteurs cités dans les deux notes précédentes, voir aussi L. J. Constantinesco, "La spécificité du droit communautaire", *Revue trimestrielle de droit européen*, 1966, p. 1-30; H. P. Ipsen, "Über Supranationalität", in *Festschrift für Ulrich Scheuner*, Berlin, 1973, P. 211-225; P. Pescatore, "International Law and Community Law - a Comparative Analysis", in *Common Market Law Review*, 1970, p. 167-183.

[89] Pour cette distinction, voir R. J. Dupuy, *Le droit international*, Paris, 1966, et R. J. Dupuy, "Communauté internationale et disparités de développement - Cours général de droit international public", *Recueil des Cours*, vol. 165, 1979, p. 9-232.

[90] M. Soerensen, "Eigene Rechtsordnungen - Skizze zu einigen systemanalytischen Betrachtungen über ein Problem der internationalen Organisation", in *Europäische Gerichtsbarkeit und nationale Verfassungsgerichtsbarkeit - Festschrift für Hans Kutscher*, Baden Baden 1981, p. 434.

[91] L'article 5 de la convention de Vienne admet que les traités ayant pour objet la constitution d'organisations internationales dérogent aux règles de la convention.

[92] A la différence, par exemple, d'accords internationaux élaborés au sein de certaines organisations internationales et qui sont soumis à ratification.

[93] Ces trois qualités caractérisent les règlements, décisions et directives du traité CEE, mais elles se retrouvent aussi, quoique moins fréquemment, dans d'autres organisations internationales.

[94] Arrêt *Costa*, attendu 2.

[95] M. Waelbroeck, *Traités internationaux et juridictions internes dans les pays du Marché communs, op. cit.*, p. 184, et suiv.; A. Koller, *op. cit.*, p. 201-209.

[96] **Voir, à ce sujet, les remarques** de A. Koller, *op. cit.*, p. 195.
[97] G. Bebr, *Development of Judicial Control of the European Communities, op. cit.*, p. 550 et suiv.
[98] D. Wyatt, *op. cit.*, p. 134: "Direct effect is not rare in international law at all, it is simply a phenomenon invariably side-stepped by international adjudicatory machinery calculated to establish State responsability."
[99] Voir *supra*, p. 269.
[100] Pour la France, voir D. Ruzie, "Les procédés de mise en vigueur des engagements internationaux pris par la France", in *Journal du droit international (Clunet)*, 1974, p. 562, à la p. 564. Pour la Suisse, voir C. Dominice, "La nature juridique des actes des organisations et des juridictions internationales et leurs effets en droit interne", in *Recueil de travaux suisses* présentés au VIIIe Congrès international de droit comparé, Basel, 1970, p. 249.
[101] Les règlements de l'ancienne Commission du Danube s'appliquaient directement aux navigants, sans la moindre intervention des États riverains; voir à ce sujet K. Von Lindeiner-Wildau, *La supranationalité en tant que principe de droit*, Leyde, 1970, p. 65 et suiv.
[102] J. H. H. Weiler, *op. cit.*, p. 69.
[103] J. A. Frowein, *op. cit.*, p. 192.
[104] R. Monaco, "Notes sur l'intégration juridique dans les Communautés européennes", in *Miscellanea Ganshof van der Meersch*, Paris-Bruxelles, 1972, II, p. 309.
[105] M. Soerensen, *op. cit.*, p. 433.
[106] Parmi les auteurs mettant en exergue les traits originaux de l'agencement juridictionnel, citons encore W. J. Ganshof Van Der Meersch, "L'ordre juridique des Communautés européennes et le droit international", in *Recueil des Cours*, vol. 148, 1975, p. 33-34, et P. Pescatore, "International Law and Community Law - a Comparative Analysis", *op. cit.*, p. 175-177. Nous avons déjà noté l'innovation du renvoi préjudiciel et l'interprétation uniforme et obligatoire qui en découle (voir *supra*, p. 14), ainsi que la spécificité du recours en manquement par rapport à l'action en responsabilité traditionnelle (voir *supra*, p. 16).
[107] En ce sens, T. Oppermann, "Die Europäische Gemeinschaft als parastaatliche Superstruktur", in *Festschrift für H. .P Ipsen*, p. 692.
[108] Par exemple, l'Empire allemand constitué en 1871.
[109] L. J. Constantinesco, "La spécificité du droit communautaire", *op. cit.*, p. 9.
[110] H. P. Ipsen, *op. cit.*, p. 220-221: "ein lediglich Vertragsformen verwendender Gesamtakt staatlicher Integrationsgewalt, der die Gemeinschaften verfaßt hat."
[111] H. .G. Schermers, *International Institutional Law*, Alphen-Rockville, 1980, p. 29 et suiv.
[112] J. H. H. Weiler, *op. cit.*, p. 54.
[113] J. H. H. Weiler, *op. cit.*, p. 367 et suiv.
[114] Dans ce sens, voir notamment M. Waelbroeck, "Contribution à l'étude de la nature juridique des Communautés européennes", in *Mélanges Henri Rolin*, Paris, 1964, p. 506-516; M. Zuleeg, "Der Bestand der Europäischen Gemeinschaft", in *Das Europa der Zweiten Generation, op. cit.*, I, p. 57 et suiv.
[115] Pour des vues d'ensemble plus fouillées sur la jurisprudence en matière de primauté, voir G. Olmi, "Les rapports entre droit communautaire et droit national dans les arrêts des juiridictions supérieures des États membres", in *Revue du Marché commun*, p. 178; G. Bebr, *Development of Judicial Control of the European Communities, op. cit.*, p. 663 et suiv. D. Carreau, "Droit communautaire et droits nationaux: concurrence ou primauté?", in *Revue trimestrielle de droit européen*, 1978, p. 405-416; J. Ferstenbert, "L'application du droit communautaire et la situation constitutionnelle du juge national", in *Revue trimestrielle de droit européen*, 1979, p. 57-72; J. V. Louis, "La primauté du droit communau-

taire", in *Les recours des individus devant les instances nationales en cas de violation du droit européen*, Bruxelles, 1978, p. 145; W. J. Ganshof Van Der Meersch, *op. cit.*, p. 216-232.

[116] J. V. Louis, *op. cit.*, p. 153.

[117] D'ailleurs, des articles similaires existent dans les constitutions de la Belgique (25 bis), du Danemark (20,1), de la France (préambule de 1946), du Luxembourg (49 bis) et des Pays-Bas (67), sans qu'ils aient joué aucun rôle dans la question de la primauté.

[118] Décision du 18 octobre 1967.

[119] Arrêt *Frontini*, n° 183, du 2 décembre 1973.

[120] On peut réapparaître ici la théorie du "transfert de souveraineté"; voir *supra*, p. 21. Elle est ainsi formulée par la cour allemande (décision du 9 juin 1971): "Si on l'interprète correctement, l'article 24,1, de la loi fondamentale signifie non seulement que le transfert des droits de souveraineté à des organisations interétatiques est licite, mais encore que les actes de leurs institutions... doivent être reconnus par le titulaire originairement exclusif du pouvoir souverain."

[121] Décision du 9 juin 1971, précitée.

[122] Respectivement, dans la décision *Internationale Handelsgesellschaft* du 29 mai 1974 et dans l'arrêt *Frontini*, précité.

[123] Les doctrines exprimées par ces trois derniers arrêts cités ont, comme on le sait, été fermement condamnées par la Cour de justice. Mais c'est là une question du plein *effet* de la primauté, et non de son *fondement*, et elle ne sera pas traitée ici.

[124] Le Conseil d'État français, quant à lui, n'a pas évolué et refuse toujours de contrôler la conformité de la loi postérieure aux traités (affaire *Syndicat général des fabricants de semoules*, arrêt du 1er mars 1968).

[125] Cour de Cassation, 24 mai 1975.

[126] Cour de Cassation belge, 27 mai 1971.

[127] P. De Visscher, *Recueil des Cours*, vol. 136, 1972, p. 42.

[128] P. Pescatore, "Rôle et chance du droit et des juges dans la construction de l'Europe".

# Reflexives Recht
# Entwicklungsmodelle des Rechts in vergleichender Perspektive

GUNTHER TEUBNER
Bremen/Florenz

Evolution des Rechts ist ein lange totgeglaubtes Thema. Beiderseits des Atlantik hat es die Rechtstheorie erst in den letzten zehn Jahren wieder belebt. Inzwischen aber hat sich in zahlreichen Varianten von Entwicklungsmodellen des Rechts ein neuer Reichtum des Evolutionismus niedergeschlagen, der angesichts der allseits bekannten "poverty of historicism" schon ein erstaunliches Phänomen darstellt[1]. Die Vielfalt der Modelle fordert geradezu den transatlantischen Theorievergleich heraus, der sich die Frage nach den Richtungsanzeigen der Evolutionsmodelle stellt: Wohin treibt die Entwicklung des modernen Rechts?

Für die europäische Diskussion verdient ein amerikanischer Beitrag besonderes Interesse, weil er in evolutionärer Perspektive klassische rechtsphilosophische Themen mit neuerem soziologischen Theoriematerial konfrontiert – die Theorie des "responsive law" von *Philip Selznick* und *Philippe Nonet.*[2]. *Nonet/Selznick* präsentieren ein Entwicklungsmodell des Rechts, in dem die innere Dynamik der Sozialinstitution Recht drei einander ablösende Stadien der Rechtsevolution hervortreibt: repressives, autonomes und responsives Recht. Jedes dieser drei – idealtypisch gemeinten – Entwicklungsstadien stellt sich als ein System von Merkmalskonstellationen dar, dessen interne Interaktionen auf eine je spezifische Krise des Rechtssystems hinsteuern, aus der sich dann das jeweils "fortgeschrittenere" Stadium entwickelt oder zumindest entwickeln kann. "Repressives Recht" als eine relativ unentwickelte Form von Recht ist im wesentlichen auf bloße Legitimation von politischer Herrschaft und Aufrechterhaltung von Ordnung ausgerichtet. Durch seine eigenen Instabilitäten bringt es eine stärker gegenüber der Politik ausdifferenzierte Form von "autonomem Recht" hervor, in der nun die Kontrolle von Macht und die Integrität des Rechts selbst die vorherrschenden Orientierungen bilden. Innere Widersprüche und Krisen dieses modernen Rechtstypus wiederum bilden die evolutionäre Chance für ein "responsives Recht", in dem Recht

als flexible, lernfähige Institution erscheint, die sensibel reagiert auf soziale Bedürfnisse und menschliche Aspirationen.

Den europäischen Gegenpart eines solchen soziologischen Entwicklungsmodells des Rechts findet man in den Evolutionstheorien von *Jürgen Habermas* einerseits und *Niklas Luhmann* andererseits. Beide thematisieren Normen- und Rechtsentwicklung im breiteren Kontext von gesellschaftlicher Evolution. In *Luhmanns* Analyse steht die Anpassung von Rechtsstrukturen an Sozialstrukturen im Vordergrund, die Frage, wie das Recht auf drei einander ablösende gesellschaftliche Differenzierungsformen reagiert: Segmentierung, Stratifizierung und funktionale Differenzierung[3]. Demgegenüber konzentriert sich *Habermas* auf die Analyse von gesellschaftlichen Orgnisationsprinzipien, die aus dem Zusammenspiel historischer Konstellationen von "sozialer Arbeit" und "kommunikativer Interaktion" resultieren[4]. Recht erscheint in *Habermas'* Modell als die institutionelle Verkörperung einer historischen Folge von präkonventionellen, konventionellen und postkonventionellen Rationalitätsstrukturen[5].

Warum ist in der Rechtstheorie heute Evolutionismus wieder en vogue? Welche Motive steuern das neuerliche Interesse an soziologischen Entwicklungsmodellen des Rechts? Vermutlich finden die im einzelnen sehr disparaten Entwicklungsmodelle ihre Einheit in einem gemeinsamen Problembezug. Sie alle suchen in evolutionärer Perspektive Antworten auf die eine Frage, nämlich wie das Recht auf die Krise seiner spezifischen modernen Rationalität – der "formalen" Rationalität – reagiert, warum diese Krise entstanden ist, welche Auswirkungen sie hat und ob eine Krisenbewältigung – oder auch nur ein Krisenmanagement – in Sicht ist.

Vor mehr als einem halben Jahrhundert hatte *Max Weber*, dessen soziologisches Interesse vornehmlich dem Spannungsverhältnis von materieller und formaler Rationalität in den verschiedensten Lebensbereichen galt, das moderne europäische Recht und – in geringerem Ausmaße – das anglo-amerikanische Recht als "formal rational" gekennzeichnet[6]. Formale Rationalität besteht danach in dem Maße, in dem professionell trainierte Juristen sich an universalistischen Normstrukturen orientieren, konkreter: in dem im Rechtsverfahren "ausschließlich eindeutige generelle Tatbestandsmerkmale materiell-rechtlich und prozessual beachtet werden"[7]. Im modernen Rechtsformalismus entspricht einer begrifflich "zunehmend logischen Sublimierung und deduktiven Strenge des Rechts" zugleich ein verfahrensmäßiges Element, eine "zunehmend rationale Technik des Rechtsganges"[8]. *Weber* analysierte bestimmte rechtliche Entwicklungsprozesse, in denen mächtige gesellschaftliche Interessen das Recht so beeinflußten, daß es von einer primär materialen, also einer inhaltlich-ethischen Orientierung zu einer formalen, also einer begrifflich abstrakten und verfahrensmäßig rationalisierten Orientierung überging.

Zugleich aber wies *Max Weber* auch nachdrücklich auf bestimmte

antiformale Tendenzen in der neueren Rechtsentwicklung hin[9]. Im Vertragsrecht etwa manifestierten sich solche Re-materialisierungen in einer "zunehmende(n) Partikularisierung des Rechts" und einer wachsenden legislativen und richterlichen Kontrolle der Vertragsinhalte. Für *Weber* bedeutete dies eine Gefährdung formaler Rationalität durch Normen anderer qualitativer Dignität: "ethische Imperative oder utilitaristische oder andere Zweckmäßigkeitsregeln oder politische Maximen, welche sowohl den Formalismus des äußeren Merkmals wie denjenigen der logischen Abstraktion durchbrechen"[10]. Nach *Max Weber* müsse die innere Qualität der hochentwickelten Rechtskultur Schaden nehmen, "wenn soziologische und ökonomische Räsonnements an die Stelle juristischer Begriffe treten"[11].

Gegenüber jenen übermächtigen Prozessen formaler Rationalisierung des Rechts jedoch erschienen für *Max Weber* diese materialen Tendenzen alles in allem nur marginal. Ganz anders nun werden in den heutigen Entwicklungstheorien diese Tendenzen eingeschätzt. "Materialisierung des Formalrechts" erscheint jetzt als der dominante neuere Entwicklungstrend, zu dessen Erklärung man evolutionistische Ansätze heranzieht[12]. In der Tat konvergieren die drei genannten neueren Entwicklungsmodelle darin, daß formale Rationalität des Rechts durch einen neuen Rationalitätstyp abgelöst wird.

Die Qualitäten dieser neuen Rechtsrationalität jedoch sind alles andere als konsentiert. *Nonet/Selznick* kennzeichnen "responsives Recht", das in der inneren Krise des Rechtsformalismus seine Entwicklungschance findet, durch eine neue "sovereignty of purpose", die sich auf sozialwissenschaftliches Wissen, auf partizipatorische Mechanismen und auf eine Repolitisierung des Rechts stützt[13]. Im Gegensatz dazu plädiert *Luhmann*, für den Krisentendenzen im formalen Recht durch mangelnde Anpassung an die funktionale Differenzierung der Gesellschaft bedingt sind, gerade umgekehrt für höhere Autonomie des Rechts, für schärfere Abstraktion der Rechtsbegriffe, für Selbstreflexion des Rechtssystems und für einen neuartigen funktionalistischen Stil der Rechtsdogmatik[14]. *Habermas* wiederum sieht die formale Rationalität des Rechts einer allgemeinen Legitimationskrise des organisierten Kapitalismus ausgesetzt, deren Überwindung die Institutionalisierung einer neuen "kommunikativen" Rationalität in den normativen, also auch in den rechtlichen Strukturen der Gesamtgesellschaft bedeuten würde[15].

Lassen sich angesichts dieses kontroversen Diskussionsstandes – so stellt sich die Frage – überhaupt plausible Aussagen über eine neue Rationalität des postmodernen Rechtes machen? Ein Weg, diese Frage zu beantworten, – und dieser wird meist beschritten – geht dahin, die konkurrierenden Rationalitätsbegriffe auf ihre jeweiligen Theorie-Hintergründe zurückzuführen und nach einer Diskussion von deren theoretischer "Richtigkeit" sich für die eine oder die andere Richtung zu entscheiden.

Man würde damit jedoch nur einen neuen Rundgang durch bekanntes Gelände mit Blick auf das beeindruckende Panorama soziologischer Großtheorien veranstalten, mit allerdings nur geringen Aussichten auf klare Richtungsanzeigen. Vielleicht sollte man deshalb seinen Weg in anderer Richtung suchen: die großen plakatierten Kontroversen meiden und nach sachlichen Übereinstimmungen und heimlichen Konvergenzen suchen, die letztlich zu weiterführenden Analysen einer neuen Rechtsrationalität hinleiten können. Das Ziel eines solchen Vorgehens ist es, dem Zwang zur Entscheidung zwischen Theorielagern auszuweichen und anscheinend unversöhnliche Standpunkte als komplementäre Perspektiven zu verstehen[16]. Konkret wird es darum gehen, die konkurrierenden Entwicklungsmodelle des Rechts in ihre Elemente zu zerlegen und diese in einer neuen Sicht wiederzuverknüpfen. In dieser Sicht konfligieren die Modelle nur vordergründig, da sie letztlich nur unterschiedliche Aspekte des gleichen Phänomens herausstellen. Möglicherweise lassen sich die drei genannten Evolutionsmodelle des Rechts im Wege wechselseitiger Anpassung, gewiß auch über notwendige Neuinterpretationen, in ein umfassenderes Modell übersetzen, das einer um die Rationalität des Rechts bemühten soziologischen Jurisprudenz weiterführende Orientierungsleitlinien vermitteln kann.

Komplementär erscheinen in unserem Fall die Modelle schon deshalb, weil *Nonet/Selznick* nur aus den "internen" Variablen des Rechtssystems und ihrer Interaktion eine Entwicklungslogik ableiten, während sowohl *Habermas* als auch *Luhmann* ihre Entwicklungstheorien explizit auf den Einfluß von "externen" gesellschaftlichen Variablen auf das Rechtssystem stützen. Deshalb mag sich die Chance eröffnen, ein umfassenderes Modell der Ko-variation *"interner" und "externer" Variablen* zu konstruieren, um damit die Entwicklungschancen eines neuartigen Rechtstypus abschätzen zu können. Welche Begrenzungen und welche Potentiale eröffnen sich für die Rationalität des Rechts, wenn man die autonome innere Entwicklungsdynamik der Rechtsstrukturen *(Nonet/Selznick)* unmittelbar konfrontiert mit Funktionsproblemen und Strukturanforderungen einer heraufziehenden post-modernen Gesellschaft *(Habermas* und *Luhmann)*?

Eine solche Re-interpretation des "responsiven Rechts" im Lichte von gesamtgesellschaftlicher Evolution soll drei Thesen plausibel machen, die im folgenden näher ausgearbeitet werden:

1. "Responsive law" als Ergebnis autonomer Rechtsentwicklung ist nicht eindimensional zu verstehen als Entwicklung einer neuen Zweckrationalität im Recht, sondern enthält in Wahrheit zwei miteinander verwobene, jedoch analytisch trennbare Entwicklungsstränge: "Formale" Rationalität wird einerseits transformiert in eine neue "materiale" Orientierung und andererseits in eine neue "reflexive" Orientierung des Rechts (Abschnitt I).

2. Jedoch stellen diese Tendenzen eher nur potentielle als aktuelle Entwicklungen dar, da die Theorie des "responsive law" hauptsächlich die

"innere Dynamik" des Rechtssystems thematisiert. Dabei bleibt die Frage offen, mit welchen externen Sozialstrukturen responsives Recht überhaupt kompatibel ist. Eine Antwort ist dann möglich, wenn man die "innere Dynamik" der Rechtsentwicklung auf "äußere gesellschaftliche Organisationsprinzipien" bezieht und diese Beziehung ihrerseits auf ihre "gesellschaftsadäquate Komplexität" untersucht (Abschnitt II).

3. Erst in der Perspektive einer Ko-variation von Rechtsstrukturen und Sozialstrukturen zeigt sich, daß die beiden verschiedenen Dimensionen von responsivem Recht durchaus unterschiedliche Verwirklichungschancen besitzen. Dies wird deutlich, wenn man "responsive law" mit Strukturproblemen der postmodernen Gesellschaft konfrontiert, nämlich mit Re-integrationsproblemen extrem ausdifferenzierter Sozialsysteme *(Luhmann)* bzw. mit Legitimationsproblemen des Interventionsstaates im organisierten Kapitalismus *(Habermas)*. Dann erreichen die Tendenzen einer "materialen " Rechtsrationalität sehr schnell die Grenzen des Wachstums, während eine "Reflexive" Orientierung durchaus Chancen besitzt, als eine sozialadäquate Antwort des responsiven Rechts zu gelten (Abschnitt III).

## I. "Toward Responsive Law": Innere Dynamik der Rechtsentwicklung

Um die verschiedenen Dimensionen des "responsiven Rechts" einschätzen zu können, scheint es notwendig, näher zu untersuchen, wie *Nonet/Selznick* ihre Hauptthese – die Transformation des Rechtsformalismus – mit ihrer Methode einer "social science strategy" verbinden. Was unterscheidet diese Strategie von herkömmlichen rechtsphilosophischen oder rechtssoziologischen Analysen? Was folgt daraus für den Begriff der Rechtsevolution? Und was ergibt sich für die beiden genannten Dimensionen von "responsive law" – für materiale und reflexive Orientierung des Rechts?

### 1. "Social Science Strategy": Jenseits von jurisprudentiellem und soziologischem Reduktionismus

*Nonet/Selznick* nehmen eine Reihe von rechtsphilosophischen Themen auf, analysieren sie jedoch mit soziologischen Mitteln[17]. In ihrer – wie sie es nennen – sozialwissenschaflichen Strategie suchen sie Recht eher mit empirischer als rein begriffanalytischer Vorgehensweise zu erfassen. Sie beziehen in einer multidimensionalen Definition des Rechtsbegriffs neben Rechtsnormen und Rechtsbegriffen explizit soziale, politische und institutionelle Elemente ein. Sie behandeln diese Elemente nicht als fixe Größen, sondern als historisch variabel in Abhängigkeit vom gesellschaftlichen Kontext. Schließlich verstehen sie spezifische Merkmalskonstellationen

des Rechts als "Systeme" von wechselseitig abhängigen Elementen, die nur eine begrenzte Anzahl von evolutionären Konfigurationen annehmen können. Mit dieser Ausweitung des Rechtsbegriffs und der expliziten Bezugnahme auf soziologische Theoriebestände gehen sie über eine traditionelle jurisprudentielle Sichtweise hinaus. Nicht die Entfaltung von Rechtsbegriffen, Rechtsprinzipien und dogmatischen Systemen steuert die Rechtsentwicklung, sondern das dynamische Zusammenspiel von sozialen Faktoren, institutionellen constraints, Organisationsstrukturen und – natürlich auch – rechtsbegrifflichen Potentialen innerhalb des Rechtssystems.

Auf der anderen Seite entgeht ein so verstandener sozialwissenschaftlicher Ansatz den Fallstricken eines soziologischen Reduktionismus. Eine häufig kritisierte Schwäche von soziologischen Theorien des Rechts ist es, in der Rückführung des Rechts auf Sozialstrukturen das Phänomen "Recht" selbst aus den Augen zu verlieren[18]. Dieser Kritik sind sowohl Basis-Überbau-Theoreme ausgesetzt, in denen Recht als bloße Widerspiegelung gesellschaftlicher Basis-Strukturen erscheint, ebenso wie "positivistische" Ansätze, die Recht auf Machtrelationen, Organisationsstrukturen, professionelle Rollen oder persönliche Attitüden reduzieren[19]. *Nonet/Selznick* vermeiden diese Art von Reduktionismus dadurch, daß sie Recht zwar soziologisch konzeptualisieren, es jedoch nicht in allgemeinen Sozialstrukturen aufgehen lassen, sondern als "autonome" Sozialinstitution begreifen, deren Entwicklung nicht bloß Reflex allgemeiner gesellschaftlicher Entwicklungen ist, sondern Resultat dynamischer Prozesse im Innern des Rechtssystems selbst.

*Autonomie der Rechtsevolution* scheint das Markenzeichen von "Law and Society in Transition" zu sein. Es handelt sich um eine Theorie von rechtsspezifischen "institutional constraints and responses", die eine bestimmte Disposition zur Veränderung damit erklärt, daß in dem einen Entwicklungsstadium sytematische Kräfte innerhalb des Rechtssystems freigesetzt werden, die ihrerseits in einem anderen Stadium charakteristische Ergebnisse produzieren[20]. Im Stadium des "repressiven Rechts" etwa ist Machtlegitimation die primäre Funktion des Rechts; und es liegt in der Logik von Legitimationsprozessen, daß sie aus sich heraus Kräfte mobilisieren, welche gerade die spezifischen Strukturen des repressiven Rechts, die sie doch erst hervorgebracht haben, unterminieren[21].

Auf diese Weise bietet *Nonet/Selznick* "social science strategy" eine interessante Alternative zu herkömmlichen jurisprudentiellen und soziologischen Interpretationen der Rechtsevolution. Ganz deutlich wird dies in der uns besonders interessierenden Deutung von Materialisierungsprozessen des Formalrechts. In der Rechtsphilosophie erklärte etwa *Radbruch* solche Einbrüche ethisch-inhaltlicher Elemente als Resultat der ewigen Spannungen zwischen Billigkeit und Rechtssicherheit[22]. Soziologische Interpretationen sehen entweder wie in der *Marx*'schen Denktradition Mate-

rialisierungstendenzen als Überbau-Reaktion auf das Auftauchen des organisierten Kapitalismus[23] oder wie *Max Weber* als Resultat des Zusammenwirkens von Marktinteressen, sozialen Klasseninteressen und internen Standesideologien der Juristen[24]. In der Analyse von *Nonet/Selznick* hingegen sind es autonome Entwicklungsprozesse im Recht selbst, die dazu führen, daß das Formalrecht bestimmte begrifflich-dogmatische Strukturen, Argumentationsweisen, Methoden der Faktenermittlung, institutionelle Konfliktlösungsverfahren und Modelle der Interessenbeteiligung hervorbringt, die dann eine solche Eigendynamik entwickeln, daß die Formalität des Rechts zerstört und ein neuer Rechtstyp produziert wird, der sich durch Zweckdenken, sozialwissenschaftliche Orientierung und politische Partizipation auszeichnet[25].

Eine solche innere Entwicklungslogik, auf der *Nonet/Selznick* insistieren, findet starken Rückhalt bei ihren europäischen Gegenspielern. *Habermas'* Rekonstruktion des Historischen Materialismus bedeutet in der Sache eine Transformation ökonomistischer Basis/Überbau-Modelle in hochabstrakte Beziehungen zwischen "sozialer Arbeit" und "kommunikativer Interaktion", die unter anderem die wichtige Modifikation enthält, daß normative Gesellschaftsstrukturen (Moral und Recht) einer autonomen "Entwicklungslogik" unterworfen sind. Die Evolution des gesellschaftlichen Moral- und Rechtsbewußtseins folge einem eigensinnigen Evolutionsmuster und könne nicht auf einen bloßen Reflex der "Entwicklungsdynamik" der sozialen und ökonomischen Basisstrukturen reduziert werden. Diese Prozesse der Rationalisierung der "Lebenswelt", zu denen die Rechts- und Moralentwicklung gehört, seien gegenüber der Komplexitätssteigerung des "Systems" autonom[26].

*Luhmann* wiederum trägt der Autonomie der Rechtsevolution dadurch Rechnung, daß er sie als Interaktion zwischen verschiedenen Evolutionsmechanismen innerhalb des Rechtssystems definiert. Normative Strukturen fungieren als Mechanismus für Variation, Verfahrensinstitutionen für Selektion, begriffliche Abstraktionen für Stabilisierung. Und deren Zusammenspiel produziert bestimmte evolutionäre Konfigurationen: archaisches, hochkulturelles und positives Recht[27]. Auch wenn die jeweiligen Entwicklungsmodelle deutliche Unterschiede aufweisen, die noch zu diskutieren sind, so sollte doch hinreichend deutlich geworden sein, daß sie in einem Punkt, dem der Autonomie der Rechtsevolution, konvergieren.

Was bedeutet es nun, Rechtsentwicklung in terms von evolutionärer Autonomie zu reformulieren? Ist dies nicht letztlich ein Rückschritt zu einem a-soziologischen Begriff der Autarkie des Rechts, der für lange Zeit eine soziologische Aufklärung der Rechtswissenschaft blockiert hat?

Autonomie des Rechts sollte nicht mit Autarkie gleichgesetzt werden, denn in allen drei Entwicklungstheorien wird das Rechtssystem explizit nicht als "geschlossenes" System in dem Sinne begriffen, daß es selbstgenügsam und unabhängig von Entwicklungen im breiteren gesellschaftli-

chen Kontext wäre. Die eigentliche Alternative zu den Interpretationen andererseits, die wir als soziologischen Reduktionismus gekennzeichnet hatten, liegt nicht im Autarkiebegriff, sondern im Begriff der Selbstprogrammierung. Ja, man sollte noch einen Schritt über *Nonet/Selznicks* Konzept von rechtlicher Autonomie hinausgehen und es als "Selbstproduktion" des Rechtssystems in seinen Elementen uminterpretieren. *Selbstreferenz* scheint der Schlüssel zum Problem zu sein, die Abhängigkeit bzw. Unabhängigkeit des Rechtssystmes von der Gesellschaft zu erfassen: Recht wandelt sich nur in Reaktion auf seine eigenen Impulse[28]. Die Rechtsordnung – Normen, Dogmatiken, Institutionen, Orgnaisation – reproduziert sich selbst in ihren Elementen, aber sie tut dies in Reaktion auf Umweltinteressen. Die Prozesse des "institutional constraint and response" finden nur zwischen internen Strukturen im Rechtssystem statt, dieser selbstprogrammierte Wandel jedoch macht das Recht sensibel für externe gesellschaftliche Anforderungen. Externe Entwicklungen werden einerseits nicht ignoriert, noch werden sie andererseits nach "Stimulus-response-Schema" direkt in interne Wirkungen umgesetzt, sondern sie werden nach Kriterien eigener Selektivität in die Rechtsstrukturen gefiltert und eingepaßt in die interne Logik normativer Entwicklung. Selbst die machtvollsten gesellschaftlichen Pressionen werden juristisch nur insoweit wahrgenommen und verarbeitet, wie sie auf den internen "Bildschirmen" der rechtlichen Wirklichkeitskonstruktionen erscheinen. In diesem Sinne "modulieren" größere gesellschaftliche Entwicklungen die Evolution des Rechts, die dabei jedoch ihrer eigensinnigen Entwicklungslogik folgt.

Selbstreferenz erlaubt also eine Sicht auf Enwicklungsprozesse des Rechtssystems, in denen dieses gleichzeitig als "offenes" und als "geschlossenes" System erscheint. Eigenständigkeit der Rechtsentwicklung und ihre Umweltabhängigkeit erscheinen nicht als Dichotomie oder als Endpunkte auf einer gleitenden Skala, auf der es mannigfaltige Abstufungen und Mischformen gibt. Vielmehr sind beide Aspekte miteinander vereinbar und gegeneinander variierbar: die Selbstbezüglichkeit rechtlicher Entwicklungen einerseits und ihre Stimulierung durch die gesellschaftliche Umwelt andererseits. Jedenfalls der erste Aspekt ist bei *Nonet/Selznick* vorzüglich herausgearbeitet; der zweite bleibt später noch zu überprüfen.

Und es ist genau der Begriff der Selbsreferenz, der es uns erlaubt, dem "distinctively legal"[29] Element auch in soziologischen Analysen des Rechts Rechnung zu tragen, ohne daß man damit eine sozialwissenschaftliche Perspektive zugunsten rechtswissenschaftlicher Autarkie aufgeben müßte. Selbstreferenz ermöglicht es uns, auch soziologisch die Unterschiede zwischen Rechtsdogmatiken und wissenschaftlichen Theorien, zwischen rechtlichen und sozialwissenschaftlichen Wirklichkeitskonstruktionen, zwischen "Rechtstatsachenforschung" und empirischer Sozialforschung, zwischen spezifisch juridischer Rationalität und der anderer Sozialsysteme nachzuvollziehen und funktional zu erklären[30]. Für unsere jetzige Analyse

dürfte jedoch das Wichtigste sein, daß Selbstreferenz in der Rechtsentwicklung unmittelbar zu einem Konzept des post-modernen Rechts führt, dessen neue Rationalität ausschließlich als Produkt einer inneren normativen Entwicklungslogik erscheint. Welches sind die Dimensionen dieser Rationalität?

## 2. Dimensionen des "responsive law": Materiale versus reflexive Orientierung

Während die aktuelle Diskussion zur Materialisierung des Formalrechts eine Fülle von Teilaspekten zu Tage gefördert hat, entwickeln *Nonet/Selznick* aus ihrer Prämisse einer autonomen Rechtsentwicklung eine systematische und kohärente Gesamtkonzeption der Rechtsmaterialisierung und ihrer Auswirkungen auf die wesentlichen Elemente des Rechtssystems. "Sovereignty of purpose" ist die neue materiale Orientierung, die das Recht in der Krise des Formalismus formt und die nicht mehr das "allgemeine Gesetz" als vorherrschende Normstruktur begünstigt, sondern offene Standards und ergebnisorientierte Direktiven. Die neue Zweckorientierung beeinflußt dogmatische Grundbegriffe ("obligation and civility") ebenso wie rechtliche Wirklichkeitskonstruktionen ("political paradigm")[31]. Sie wirkt sich weiter auf die Rechtsmethodik aus: klassische Methoden der Rechtsfindung werden überlagert von Methoden einer "social policy analysis". Verfahren und Beteiligungsmodi im klassischen Zwei-Parteien-Prozeß wandeln sich zu eher politischen Konfliktlösungsmustern ("legal pluralism")[32]. Darüber hinaus erfordert die Materialisierung des Rechts völlig neue institutionelle und organisatorische Strukturen. Verlangt ist "regulation, not adjudication", die von nicht-hierarchischen "post-bureaucratic organization" implementiert wird. Schließlich müssen die geschlossenen Außengrenzen des "autonomen" Rechts in Richtung auf Politik und Gesellschaft geöffnet werden: Materiale Rationalität erfordert eine "integration of legal and moral judgment and of legal and political participation"[33].

Es lassen sich jedoch im "responsiven Recht" auch Elemente ausmachen, die in einer materialen Orientierung des Rechts eigentlich Fremdkörper darstellen. Obwohl *Nonet/Selznick* auch diese Elemente ausdrücklich unter "substantive justice" subsumieren, im Gegensatz zu "procedural fairness" im autonomen Recht und zu "raison-d'état" im repressiven Recht[34], ist zu vermuten, daß sie der Logik einer ganz anderen Rationalität gehorchen.

Deutlich wird dies am Beispiel von "institutional design and institutional diagnosis", das im Denken von *Nonet/Selznick* einen zentralen Platz einnimmt[35]. Hier ist die Aufmerksamkeit des Rechts darauf gerichtet, soziale Institutionen zu schaffen, zu gestalten, zu korrigieren und zu redefinieren, die als selbstregulierende Systeme funktionieren. Rechtsnor-

men werden in diesem Zusammenhang nicht im Hinblick darauf entworfen, gesellschaftliche Prozesse konkret in Einzelheiten zu steuern, sondern abstrakter, um strukturelle Kongruenzen zwischen gesellschaftlichen Ordnungen und rechtlichen Normen herzustellen. Statt auf inhaltliche Verhaltenssteuerung sind diese Normen auf Organisation, Verfahren und Kompetenzen ausgerichtet. Das Recht übernimmt hier nicht die Verantwortung für bestimmte soziale Ergebnisse, sondern zieht sich darauf zurück, selbstregulatorische Mechanismen zu regulieren, wie etwa Verhandlung, Dezentralisierung, Planung, organisierter Konflikt. Während eine materiale Rationalität eine umfassende Regulierung von Sozialprozessen erfordert, zielt "institutional design" auf "enablement and facilitation".

Gleiches läßt sich von anderen Elementen des "responsiven Rechts" sagen. Die geforderte Politisierung des Rechts etwa, die sich in neuen Formen politischer Partizipation im Rechtssystem (social advocacy, class action, Repräsentation von Gruppeninteressen im Prozeß) äußert[36], hat die Wirkung, sowohl neue gesellschaftliche Konflikte zu internalisieren als auch neue gesellschaftliche Interessen ins Recht zu integrieren. Dies geschieht ohne direktes Engagement für konkrete soziale Ergebnisse, was für eine materiale Rationalität aber erforderlich wäre. Weiterhin ist es nicht auszuschließen, daß das Eindringen der Sozialwissenschaften ins Recht gar nicht so sehr mit Ergebnis- und Folgenorientierung, mit sozio-technischer Steuerung gesellschaftlicher Prozesse, zu tun hat, wie es gemeinhin insbesondere mit "social-policy"-Ansätzen assoziiert wird[37]. Sozialwissenschaftliches Denken mag durchaus anderswo für das Recht relevant werden, nämlich für eine Umstellung abstrakter Modell-Konstruktionen, für einen Wandel in begrifflichen Grundstrukturen, für einen andersartigen Stil der Argumentation. Solche Wandlungen in den rechtlichen Fundamenten haben nur noch wenig zu tun mit der "wissenschaftlichen" Produktion inhaltlicher Ergebnisse im Sinne materialer Rationalität.

Diese Überlegungen legen es nahe, eine klare analytische Unterscheidung in das Konzept des "responsive law" einzuführen. In der Transformation des Rechtsformalismus lassen sich – wie schon angedeutet – zwei Trends diagnostizieren: Auf dem Wege zum responsiven Recht wird die "formale" Rationalität des Rechts einerseits durch eine "materiale" Orientierung ersetzt, andererseits durch eine "reflexive" Orientierung. Dies bedarf einiger begrifflicher Klärungen.

Als hilfreich erwiesen sich hierzu Differenzierungen, wie sie *Habermas* eingeführt hat, um die Rationalität des modernen (formalen) Rechts näher zu analysieren[38]. Diese Differenzierungen sollte man jedoch nicht nur in bezug auf formales Recht anwenden, sondern weitergehend auch für den materialen und den reflexiven Rechtstyp fruchtbar machen. Man kann damit die Unterschiede der genannten Rechtstypen in den folgenden Dimensionen erfassen: Sie unterscheiden sich spezifisch (1) in der internen

Systematisierung des Rechtsstoffes *(interne Rationalität)*, (2) im spezifischen Modus der Rechtfertigung von Geltungsansprüchen *(Normrationalität)* und (3) in ihrem Beitrag zur Bestandserhaltung der Gesellschaft *(Systemrationalität)*.

*Formale Rationalität* des modernen Rechts ist dann durch drei historisch miteinander verbundene Elemente definiert: (1) In seiner "internen Rationalität", die sich auf die Begriffsstrukturen des Rechts bezieht, ist Recht formal rational in dem Maße, in dem es an Kriterien analytischer Begrifflichkeit, deduktiver Strenge und Normanwendung im Sinne eindeutiger Tatbestandsorientierung ausgerichtet ist. Im formalen Recht wenden Rechtsexperten formal-operationales Denken auf ihr professionelles Wissen in universalistischer Weise an[39]. (2) "Normrationalität" hingegen bezieht sich auf die grundlegenden Prinzipien, welche die Art und Weise rechtfertigen, in der Rechtsnormen menschliches Handeln ordnen sollen. In dieser Dimension bedeutet "Formalität" des Rechts etwas ganz anderes, nämlich, daß das Recht klar beschränkt sein soll auf die Ausgrenzung abstrakter Handlungssphären für die autonome Verfolgung privater Interessen. Recht garantiert nur eine "formale" Rahmenordnung, während die "materialen" Entscheidungen von privaten Akteuren gefällt werden. Formalrecht in diesem Sinne ermöglicht private Ordnungen, es beruht auf der Indifferenz des Gesetzgebers gegenüber der Frage, welche alternativen Beziehungen die privaten Akteure eingehen sollen[40]. Die Ordnungsleitung des Formalrechts ist nur negativ im Sinne einer Begrenzung prinzipiell gewährter subjektiver Rechte; Formalrecht ordnet keine konkreten Pflichten und inhaltliche Verhaltensregeln an. "An die Stelle vorgegebener und beurteilter Symbiosen von Rechten und Pflichten tritt die soziale Ermächtigung zum Handeln"[41]. Seine komplementären Elemente sind – in *Habermas'* Terminologie – Konventionalität, Legalismus und Universalität[42]. (3) "Systemrationalität" schließlich bezieht sich auf die Kapazität der Rechtsordnung, auf Kontrollprobleme der Gesellschaft zu antworten. Rational in diesem Sinne ist das Formalrecht, insoweit es die normativen Imperative einer entwickelten Marktgesellschaft erfüllt, insoweit es die "Erschließung, Mobilisierung und zweckmäßige Allokation von natürlichen Ressourcen und von Arbeitskraft fördert" *(Habermas)*. Die "Semantik der Dezentralisation" als Kennzeichen des Formalrechts, das in Form von subjektiven Rechten Autonomiespielräume ausgrenzt, stellt die Systemrationalität dieses Rechtstyps in bezug auf eine ausdifferenzierte Wirtschaft her[43].

Unser Vorschlag geht nun dahin, diese drei Dimensionen von Rationalität auch für den "materialen" und den "reflexiven" Rechtstyp fruchtbar zu machen. *Materiale Rationalität* hängt historisch mit der zunehmenden "Verrechtlichung" vormals informeller sozialer Prozesse zusammen. Sie wird gemeinhin assoziiert mit dem Wachstum des Sozialstaats und mit staatlicher Intervention in organisierte Marktstrukturen[44]. In diesen Ent-

wicklungen verliert das Recht seinen formalen Charakter in allen drei Dimensionen. (1) Die "interne Rationalität" wandelt sich, insofern die vorherrschende Regelorientierung von einer zunehmenden Zweckorientierung überlagert wird. Statt einer strikten Anwendung präzis definierter Rechtsnormen (Konditionalprogramme) verwalten jetzt Rechtsexperten eher offen definierte Standards und vage Generalklauseln (Zweckprogramme).Die daraus resultierenden Strukturprobleme für die Rechtsdogmatik sind bekannt[45]. (2) In seiner "Normrationalität" verschiebt sich der Fokus des Rechts von Autonomie auf Regulierung. Statt Freiheitssphären für autonomes privates Handeln zu garantieren, tendiert materiales Recht dazu, soziales Verhalten direkt und ergebnisorientiert zu steuern. Nun definiert die Rechtsordnung selbst konkrete Pflichten und inhaltliche Verhaltensregeln. Gleichzeitig verliert es seine universalistische Orientierung, welche das Rechtssubjekt von sämtlichen Statusbeziehungen abstrahierte, und tendiert zu einem neuen Partikularismus, einer neuerlichen Orientierung der Rechtsnormen an gesellschaftlichen Rollenzusammenhängen[46]. (3) Materiales Recht gewinnt "Systemrationalität" in dem Maße, in dem es dogmatische Figuren und Verfahrensmodi herausbildet, die für die politischen Interventionen des modernen Wohlfahrtsstaates adäquate Rechtsformen liefern. Materiales Recht ist instrumentalisierbar für die Zwecke des politischen Systems, das nun für gesellschaftliche Prozesse die Verantwortung übernimmt, und zwar sowohl in der Definition von Zielen, in der Auswahl normativer Mittel, im Anordnen konkreter Verhaltensprogramme als auch in der Implementierung der Normen. Erst durch seine "Materialisierung" wird Recht zum wichtigsten Steuerungsmedium des Wohlfahrtsstaates.

*Reflexive Rationalität* stellt demgegenüber eine weitaus weniger wohldefinierte Orientierung des Rechts dar, was mit ihrem noch relativ unentwickelten Status zusammenhängt. Dieser Rationalitätstyp tauchte erst in neuerer Zeit auf, im Grunde erst in der sich abzeichnenden Krise des Wohlfahrtsstaates, als eine noch wenig geklärte Alternative zu den zur Zeit recht dominanten regressiven Tendenzen der Re-formalisierung des materialen Rechts. Mit staatsinterventionistischen Konzepten teilt reflexives Recht das Programm eines Rechtsaktivismus, das in soziale Prozesse kompensatorisch zu intervenieren sucht. Jedoch zieht es sich aus der vollen Verantwortung für konkrete soziale Ergebnisse zugunsten einer abstrakteren Steuerung zurück. (1) Die "interne Rationalität" eines reflexiven Rechts liegt jenseits der Alternative von Konditionalprogrammen versus Zweckprogrammen. Sie baut weder auf ein System von präzis definierten Normen und Rechtsbegriffen, noch auf die Zweck-Mittel-Logik von materialen Rechtsprogrammen. Reflexives Recht tendiert eher zu abstrakteren prozeduralen Programmen, die sich auf die Meta-Ebene der Regulierung von Prozessen, von Organisationsstrukturen, auf die Verteilung und Neudefinition von Steuerungsrechten und von Entscheidungskompetenzen

zurückziehen. Ein solcher neuer "Prozeduralismus" kann auf ganz verschiedenen Rechtsgebieten beobachtet werden, als eine sich entwickelnde Alternative sowohl zu formaler als auch materialer Rechtsrationalität. *Wiethölter* etwa analysiert auf dem Gebiet des Sonderprivatrechts diese neuere Tendenz in drei Entwicklungsstufen: Formalisierung, Materialisierung, Prozeduralisierung, wobei er die letztere sozusagen als steckengebliebene Entwicklung auf dem Wege zu einer Materialisierung begreift[47]. *Eike Schmidt* spricht im Vertragsrecht von einer neuen "Sozialautonomie", welche die Defizienzen klassischer Privatautonomie gerade nicht durch Materialisierung, sondern durch Prozeduralisierung überwinden soll[48]. Auf dem Gebiet des Verbraucherrechts analysiert *Joerges* ein "Entdeckungsverfahren Praxis" jenseits der materialisierten "Verwaltung des Privatrechts"[49]. Es zielt auf die Koordination von "Politiken unterschiedlicher gesellschaftlicher Akteure. Auch wenn hier eine anders nuancierte Interpretation dieses neuen Rechtstyps, im Zusammenhang mit einer Selbstreflexion des Rechtssystems, versucht werden soll, so kann doch als konsentiert unterstellt werden, daß die prozedurale Orientierung des Rechts indirektere, abstraktere Mittel sozialer Kontrolle bevorzugt: reflexive Mechanismen, die anstelle einer inhaltlichen Entscheidung selbst nur über die organisationalen und prozeduralen Entscheidungsprämissen entscheiden[50]. (2) Die "Normrationalität" des reflexiven Rechtstyps weist in der Tat deutliche Parallelen zu neo-liberalen Rechtskonzeptionen auf. Indem es soziale Autonomie mit Rechtsmitteln schafft, fördert und neu definiert, setzt es letztlich auf "invisible-hand"- Mechanismen. Im direkten Gegensatz zu diesen Konzepten beschränkt es sich aber nicht darauf, sich an "natürliche soziale Ordnungen" anzupassen, bzw. deren Wiederherstellung mit der "sichtbaren Hand des Rechts" zu betreiben. Reflexives Recht zielt auf "regulierte Autonomie", es fördert aktiv selbst-regulierende "lernende" Sozialsysteme und versucht zugleich, deren Defizienzen mit kompensatorischen Korrekturen abzubauen, und dies alles mit abstrakten Steuerungsmitteln von Organisation und Verfahren. Es anerkennt nicht bloß vorgegebene subjektive Rechte, sondern steuert durch die Neudefinition und Neuverteilung von fungiblen property rights. (3) Die "Systemrationalität" eines reflexiven Rechtstyps schließlich ist an einem Zentralproblem hochdifferenzierter Gesellschaften ausgerichtet, dem Problem der Integration autonomer Subsysteme. Es geht jedoch nicht in klassischer Perspektive um "normative" Integration über die Projektion von gemeinsamen Normen und Werten für alle Teilsysteme der Gesellschaft, sondern um die Definition rechtlicher Strukturprämissen für eine dezentralisierte Integration der Gesellschaft. Die Rolle reflexiven Rechts besteht dann darin, integrative Mechanismen für Verfahren und Organisation innerhalb der betroffenen Teilsysteme selbst bereitzustellen, ihnen eine "Sozialverfassung" zu geben, die ihre Eigengesetzlichkeiten respektiert, ihnen aber zugleich gesellschaftliche Restriktionen auferlegt.

Warum "reflexive" Rationalität? Der theoretische Kontext, auf den sich die Terminologie bezieht, wird im Teil III noch systematisch behandelt. Um die Begriffswahl vorläufig zu rechtfertigen, mag der Hinweis genügen, daß der Ausdruck "reflexiv" in unserem Zusammenhang drei Phänomene zusammenfassend kennzeichnet. Er bezeichnet erstens Selbstidentifikationsprozesse im Rechtssystem als solchem; er verweist zweitens auf die unterstützende Rolle des Rechts in Selbstidentifikationsprozessen anderer Sozialsysteme; er bezeichnet schließlich drittens die selbstbezüglichen normativen Mechanismen, deren sich die Rechtsordnung dabei bedient.

Als Illustration der vorgestellten abstrakten Begriffsbildungen soll wiederum das schon erwähnte Vertragsrecht dienen. In der rechtlichen Transformation der Sozialbeziehung "Vertrag" betont jeder der genannten Rationalitätstypen des Rechts einen je spezifischen Aspekt. Formale Bedingungen der Willenübereinstimmung abstrakter Rechtssubjekte, die generellen Wirksamkeitsvoraussetzungen des Vertragsabschlusses, stehen bekanntlich im Zentrum des klassischen Formalrechts. Während Materialisierungsprozesse zu einer verstärkten richterlichen und gesetzgeberischen Inhaltskontrolle von Verträgen führen, würde ein reflexiver Rechtstyp seine Aufmerksamkeit nicht nur auf die prozeduralen und organisationalen Bedingungen der Machtneutralisierung im Verhandlungsprozeß richten, sondern insbesondere auf vertragliche Mechanismen "öffentlicher Verantwortung", die das kooperative System sensibel gegenüber seinen sozialen Effekten machen würden. Daß hier die rechtsförmige Ausgestaltung des Tarifvertragssystems als recht fortgeschrittene Form des Vertragsrechts besonders einschlägig ist, braucht nicht erst erwähnt zu werden.

"Externe Dezentralisierung" mag als ein weiteres Beispiel für reflexive Entwicklungschancen des Rechts stehen[51]. In der aktuellen Diskussion über "Verrechtlichung" und die "Grenzen des Sozialstaats" ist immer wieder herausgearbeitet worden, welche disfunktionale Konsequenzen es hat, wenn sozialstaatliche Steuerung Lebensbereiche "verrechtlicht", deren strukturelle Eigenarten sich den Steuerungsmedien Geld und Recht widersetzen. Die Materialisierung des Rechts findet an solchen Lebensbereichen, die vormals nur informell oder allenfalls über Formalrecht strukturiert waren, ihre natürlichen Grenzen. Die bekannten neo-konservativen Lösungsformeln heißen: De-regulierung, Re-formalisierung, Privatisierung. Als Alternative wurde das Konzept der "externen Dezentralisierung" entwickelt, wonach öffentliche Aufgaben zwar auf halbstaatliche oder private Institutionen übertragen werden, aber dennoch in dem Sinne öffentlich bleiben, daß sie politisch verantwortet werden müssen. Delegationen auf "korporatistische" Gremien, halbstaatliche Verbände, auf die Tarifpartner oder auf kommunale gesellschaftliche Organisationen sind Beispiele für solche "externe Dezentralisierung". Die politische Gesamt-

verantwortung wird auf eine rechtliche Rahmenregelung, auf eine "Sozialverfassung", auf Binnenkonstitutionalisierung, also auf Partizipationsrechte und Entscheidungsverfahren, zurückgenommen mit der jederzeitigen Möglichkeit zu Korrektur und Revision der institutionellen Arrangements, wenn sich die Ergebnisse solcher gesellschaftlichen Lernprozesse als politisch unzuträglich erweisen sollten.

Um es klarzustellen: Im Konzept des "responsive law" diskutieren *Nonet/Selznick* durchaus Elemente sowohl von materialer als auch von reflexiver Rationalität des Rechts, ohne jedoch diese systematisch zu unterscheiden. Was macht die Unterscheidung aber so relevant? Reflexive Rechtsrationalität benötigt ganz andere institutionelle Ordnungen, ganz andere Wirklichkeitsmodelle und ganz andere normative Strukturen als eine materiale Orientierung. Wichtiger ist noch ein zweiter Gesichtspunkt. Erst die systematische Unterscheidung beider Rechtstypen erlaubt es, die Verwirklichungschancen des "responsive law" im gesellschaftlichen Kontext abzuschätzen, nämlich dann, wenn man die Strukturmerkmale beider Rechtstypen systematisch auf die Funktionsprobleme ihrer gesellschaftlichen Umwelt bezieht.

An dieser Stelle stoßen wir nun auf eine empfindliche Schwäche in der Konzeption von *Nonet/Selznick*. Gerade ihr Begriff von Autonomie der Rechtsentwicklung erweist sich nun als problematisch. Könnte es sein, daß für den analytischen Gewinn dieses Begriffes, der das Verständnis der inneren Entwicklungsdynamik außerordentlich fördert, ein zu hoher Preis gezahlt werden muß? Der Preis, den *Nonet/Selznick* für den Begriff der autonomen Rechtsentwicklung entrichten müssen, scheint darin zu bestehen, daß sie die Interdependenzen von rechtlichen und gesellschaftlichen Entwicklungsprozessen umfaßt, dabei aber die systematischen Einflüsse der gesellschaftlichen Umwelt, die innerrechtliche selbstreferentielle Prozesse stimulieren, vernachlässigt. Deshalb können sie ohne weiteres im Begriff des "responsive law" materiale und reflexive Tendenzen zusammenfassen, da sich für beide aus der Sicht des Rechtssystems keine Unterschiede in ihrer Realisierungsmöglichkeit zeigen. Will man hingegen die gesellschaftlichen Wirkungschancen genauer analysieren, so muß man "interne" und "externe" Variablen in einem Modell der Ko-variation von Rechtsstrukturen und Sozialstrukturen aufeinander beziehen. Und in einer solchen Perspektive soll im folgenden unserer These begründet werden, daß nämlich soziale Entwicklungen außerhalb des Rechtssystems das Potential "materialer Rechtsrationalität drastisch beschränken, während die gleichen Entwicklungen die Wirkungschancen des "reflexiven" Rechtstyps eher systematisch fördern. Oder um es in *Nonet/Selznicks* eigener Terminologie auszudrücken: Sollte man eher im "reflexiven" als im "materialen Rechtstyp die notwendige "conceptual readiness" finden, mit der das responsive Recht auf die "opportunity structures" der gesellschaftlichen Entwicklung reagieren kann?[52]

*Typen und Dimensionen der Rationalität des Rechts*

| Typen / Dimensionen | Formal | Material | Reflexiv |
|---|---|---|---|
| Interne Rationalität | Regelorientierung: Konditionalprogramme, präzise Definition von Tatbestandsmerkmalen und Rechtsfolgen; dogmatische Systembildung | Zweckorientierung: Zweckprogramme, vage Standards und Generalklauseln; offene (topische) Argumentation | Verfahrensorientierung: Reflexive Normierung, Organisations-, Verfahrens-, Kompetenznormen; Logik selbstregulierender Systeme |
| Normrationalität | proskriptiv: grenzt Autonomiesphären für private Akteure aus | präskriptiv: reguliert direkt und ergebnisorientiert soziale Prozesse | fazilitativ: reguliert indirekt und abstrakt soziale Selbstregulierung |
| Systemrationalität | rechtliche Strukturvoraussetzungen für Mobilisierung und Allokation von Ressourcen in einer entwickelten Marktgesellschaft | Rechtsinstrumente für politische Interventionen des Wohlfahrtsstaates: Korrektur von Marktdefiziten und kompensatorische soziale Maßnahmen | integrative Funktion: rechtliche Verfahrens- und Organisationsprämissen für Reflexionsprozesse innerhalb sozialer Subsystem |

## II. "Law and Society in Transition": Recht ohne Gesellschaft?

### 1. Ein Modell inneren Wachstums

Das Interesse gilt jetzt der Beziehung zwischen Gesellschaft und Recht in den Evolutionsmodellen des Rechts. Bei *Nonet/Selznick* erscheint diese Beziehung nur marginal. Ganz im Vordergrund steht die innere Dynamik. Sie setzt systematische Kräfte frei, die neuartige Konfigurationen im Rechtssystem produzieren. Äußere gesellschaftliche Entwicklungen erscheinen bloß als historisch kontingente Ereignisse, die Entwicklungen innerhalb der Sozialinstitution Recht blockieren oder fördern mögen, nicht aber die Merkmalskonstellationen des Rechtssystems systematisch beeinflussen[53]. Zwar betonen die Autoren durchweg, daß die Variationen in den vielen Dimensionen des Rechts auch vom gesellschaftlichen Kontext abhängen, jedoch versäumen sie es, ihr Entwicklungsmodell des Rechts mit einem ähnlich sorgfältig ausgearbeiteten Entwicklungsmodell breiterer gesellschaftlicher Strukturen zu kombinieren.

Man wird ihnen auch einräumen müssen, daß die gesellschaftlichen Bezüge der Rechtsentwicklung in ihrer konkreten Ausarbeitung der drei Rechtstypen weitaus vielfältiger und reicher sind als ihr formales "developmental model" vermuten läßt. Insbesondere sind die Beziehungen zwischen der rechtlichen und der politischen Ordnung als ein spannungsreicher und prekärer Entwicklungsprozeß dargestellt, so daß man ihre Schrift als eine Spielart einer politischen Rechtstheorie lesen kann. Dennoch müssen sie sich die Kritik gefallen lassen, daß in der Theorie des responsive law die Gesamtgesellschaft allenfalls über Politik repräsentiert ist, was eine drastische Verengung bedeutet. Und zweitens, daß der Begriff einer Entwicklungslogik ausschließlich für das Rechtssystem entfaltet wird, nicht aber für Politik oder sonstige gesellschaftliche Teilsysteme und schon gar nicht für die Gesellschaft als Ganzes.

*Nonet/Selznicks* Ansatz läßt sich am besten als ein Modell inneren Wachstums kennzeichnen, in dem die Rolle gesellschaftlicher Strukturen darauf beschränkt ist, für vor-programmierte Wachstumsprozesse im Recht mehr oder weniger günstige Umweltbedingungen bereitzustellen. Wachstum oder Niedergang aber sind im Rechtssystem selbst angelegt. Insofern bietet uns "responsive law" das Bild eines Rechts ohne Gesellschaft. Äußere gesellschaftliche Kräfte dienen allenfalls dazu, ein bereitstehendes Potential zu aktualisieren, die Stabilität eines rechtlichen Entwicklungsniveaus zu sichern bzw. die Wahrscheinlichkeit von Fortschritt oder Regression zu beeinflussen.

Dieses Wachstumsmodell des Rechts mit nur minimalen Bezügen auf gesellschaftliche Evolution läßt einige Fragen offen. Welche Mechanismen besorgen eigentlich die gesellschaftliche Aktualisierung des Entwicklungspotentials im Recht und welche analytischen Werkzeuge sollte man auf

sie anwenden? Muß man sie sich als Druck gesellschaftlicher Interessen vorstellen, deren Wirkung auf das Recht mit den Mitteln einer Kausalanalyse erhellt werden kann? Oder ist eine Zweck-Mittel-Analyse eher angemessen, mit der sich erfassen läßt, wie individuelle und kollektive Akteure Rechtsinstitutionen für eigene Zwecke instrumentalisieren? Eine dritte Möglichkeit wäre eine funktionalistische Analyse, die Beziehungen zwischen externen gesellschaftlichen Funktionen und internen Rechtsstrukturen aufdeckt. All diese Fragen bleiben bei *Nonet/Selznick* offen.

Eine grundsätzlichere Kritik noch ist anzubringen gegenüber der bloßen Statistenrolle, die gesellschaftliche Kräfte in *Nonet/Selznicks* Inszenierung spielen dürfen. Gehen bei dieser Regie nicht wichtige Botschaften der "Klassiker" der Rechtssoziologie verloren? *Durkheims* "restitutives" Recht, das die organische Solidarität moderner Gesellschaften repräsentiert, ist verständlich nur vor der Staffage einer bestimmten Sozialstruktur, nämlich vor gesellschaftlicher Teilung der Arbeit, die an die Stelle einer segmentierten Organisation der Gesellschaft mit nur mechanischer Solidarität getreten ist[54]. Ebenso lassen sich bei *Max Weber* Rationalisierungsprozesse im Recht nur im Zusammenspiel mit ähnlichen Erscheinungen in Wirtschaft, Politik und Wissenschaft darstellen[55]. Kann man solche Einsichten über systematische gesellschaftlich-rechtliche Interaktionen – so kritikbedürftig sie im einzelnen auch sein mögen – reduzieren auf die Denkfigur von Potential und Aktualisierung? Noch radikaler gefragt: Macht es angesichts durchgängiger sozialer Interdependenzen überhaupt einen Sinn, Entwicklungsmodelle für ein spezifisches Sozialsystem, in unserem Falle Recht, zu konstruieren, ohne gleichzeitig eine Theorie gesamtgesellschaftlicher Entwicklung wenigstens in den Grundzügen vorzulegen?

Ein dritter Komplex von Fragen stellt sich in bezug auf den von *Nonet/Selznick* vorgestellten Krisenbegriff. In jeder der drei Entwicklungsphasen soll eine spezifische Krise des Rechts entstehen, die Entwicklungschancen für einen neuen Rechtstyp auftauchen läßt. Doch bleibt die Frage nach der Natur dieser Krise. Stellen sich in diesen Krisen Probleme der "Systemintegration" in dem Sinne, daß die jeweiligen Rechtsstrukturen einen geringeren Möglichkeitsspielraum eröffnen, als für die Kontrollprobleme der Gesellschaft erforderlich wäre[56]? *Nonet/Selznicks* Ausführungen zu größeren "Problemlösungskapazitäten" des Rechts, zur "Ökonomie der Macht" und zur "prekären Balance" im responsiven Recht würden durchaus in einen solchen funktionalistischen Analyserahmen passen[57]. Andererseits thematisieren sie jedoch die Probleme des repressiven wie des autonomen Rechts als spezifische Legitimationskrisen, was auf Fragen der "Sozialintegration" verweisen würde, auf defiziente soziale Identität und auf normative Desintegration. Möglicherweise würden *Nonet/Selznick* sich dagegen verwahren, vor eine solche klar geschnittene Alternative gestellt zu werden, da ihr Krisenkonzept Elemente beider enthält. Da wir jedoch

Sozialintegration und Systemintegration nicht gleichsetzen können, müssen wir dann nicht jedenfalls innerhalb eines umfassenderen Krisenkonzepts deren interne Beziehungen klären? Denn eine Rationalisätskrise des Rechts, die ihre sozialtechnischen Kapazitäten betrifft, hat andere Ursachen, Auswirkungen und Lösungsmöglichkeiten als eine Legitimationskrise. Wie etwa muß man die Krise des autonomen Rechts, die Krise formaler Rationalität in *Max Webers* Sinne, verstehen? Als Kontrolldefizit der abstrakten, allgemeinen und formalen Rechtsnormen[58]? Oder als Legitimationsproblem des Rechtsformalismus, der sich Fragen substantieller Gerechtigkeit nicht stellt[59]? Eine Klärung dieser Fragen hätte beträchtliche Auswirkungen auf das Konzept eines responsiven Rechts. Die Herausforderungen, auf die responsives Recht antwortet, könnten präziser analysiert werden, entweder als interne Legitimationsprobleme oder als externe Rationalitätsprobleme des Rechts. Mehr noch, die Antworten des responsiven Rechts würden entsprechend unterschiedlich ausfallen: Verstärkung des Steuerungspotentials einerseits, Ermöglichung von sozialen Identifikationsprozessen andererseits.

Alle drei Fragenkomplexe gehen über eine bloße Klärung der Modellprämissen von *Nonet/Selznick* hinaus. Sie tendieren dahin, die Reichweite des Entwicklungsmodells in einer bestimmten Richtung zu vergrößern: von internem Wachstum des Rechts zu gesellschaftlich-rechtlicher Kovariation. Eine solche Erweiterung mag die Chance eröffnen, *Nonet/Selznicks* Analysen der inneren Rechtsdynamik zu nutzen, aber gleichzeitig die klassischen Intentionen einer umfassenden Gesellschaftsanalyse des Rechts beizubehalten.

## 2. Alternative Modelle: Ko-variation von Rechtsstrukturen und Sozialstrukturen

Für diese Zwecke mag es fruchtbar sein, zwei zentrale Begriffe heranzuziehen, die die Ko-variation von Recht und Gesellschaft betreffen: den Begriff des "Organisationsprinzips der Gesellschaft" *(Habermas)* und den der "sozialadäquaten Komplexität des Rechts" *(Luhmann)*. In beiden Begriffen ist Rechtsentwicklung als autonome angelegt, zugleich aber ist ein je spezifischer Zusammenhang von Recht und Gesellschaft systematisch ausgearbeitet.

*Habermas* argumentiert, daß nachdarwinianische Theorien der Gesellschaftsevolution, die sich auf das Zusammenspiel der drei bekannten Evolutionsmechanismen stützen (Variation, Selektion, Stabilisierung), nicht in der Lage sind, die Identität von Entwicklungsstufen und deren Lernpotential zu analysieren. Beides könne erst von einem Begriff gesellschaftlicher Bewußtseinsstrukturen erfaßt werden[60]. In der Konsequenz kontrastiert *Habermas* die "Entwicklungsdynamik", die aus dem Zusammenspiel der Evolutionsmechanismen in der Gesellschaft entsteht, mit

einer autonomen "Entwicklungslogik" normativer Strukturen (Moral- und Rechtsbewußtsein), die ein rational nachkonstruierbares Muster evolutionärer Sequenzen hervorbringen soll. *Habermas* benutzt Modelle der moralischen Ich-Entwicklung in der Tradition von *Piaget* und *Kohlberg* und übersetzt sie aus individuellem in gesellschaftlichen Kontext[61]. Aus dem Zusammenspiel der "Entwicklungslogik" normativer Strukturen und der "Entwicklungsdynamik" gesellschaftlicher Basisstrukturen resultieren dann bestimmte "Organisationsprinzipien" der Gesellschaft, die ihrerseits eine logische Folge strukturierter Ganzheiten bilden. Der Entwicklungsprozeß dieser Organisationsprinzipien ist gekennzeichnet durch Irreversibilität, strukturierte Hierarchie und evolutionäre Gerichtetheit[62].

In *Habermas'* Evolutionstheorie definiert das jeweilige Organisationsprinzip – und damit das Recht als Teil seiner institutionellen Verkörperung[63] – das Lernniveau einer gegebenen Gesellschaft[64]. Es bestimmt die Variationsbreite für Sozialintegration (soziale Identität, Wertkonsens) ebenso wie für Systemintegration (Kontrollkapazität der Gesellschaft). Organisationsprinzipien sind nach *Habermas* das Produkt eines doppelten Lernprozesses, der angemessen nur durch die Kombination zweier analytischer Modelle erfaßt werden kann. Ein funktionalistisches System/Umwelt-Modell erweist sich als fruchtbar bei der Analyse der gesellschaftlichen Kontrollkapazität eines Organisationsprinzips. Im Falle einer Systemkrise jedoch, wenn Entwicklungen in der Sozialsphäre Probleme schaffen, die die Kontrollkapazität eines gegebenen Organisationsprinzips übersteigen, setzen neue Lernprozesse innerhalb der kulturell-normativen Sphäre ein, zu deren Erklärung dann aber ein anderes Modell herangezogen werden muß, das Modell der "rationalen Rekonstruktion". Mit diesem Modell soll es möglich sein, normativen Mustern in ihrer eigentümlichen Entwicklungslogik nachzuspüren, die letztlich kulturelle Innovationen und die Herausbildung neuartiger Organisationsprinzipien ermöglicht. *Habermas* selbst kennzeichnet die systematisch nachkonstruierbaren Muster der Entwicklung normativer Strukturen folgendermaßen[65]:

> "Diese Strukturmuster beschreiben eine den kulturellen Überlieferungen und dem Institutionenwandel innewohnende *Entwicklungslogik*. Diese sagt nichts über die Entwicklungsmechanismen; sie sagt nur etwas über den Variationsspielraum, innerhalb dessen kulturelle Werte, Moralvorstellungen, Normen usw. auf einem gegebenen Organisationsniveau der Gesellschaft verändert werden und verschiedene historische Ausprägungen finden können. In seiner *Entwicklungsdynamik* bleibt dieser Wandel normativer Strukturen abhängig von den evolutionären Herausforderungen ungelöster, ökonomisch bedingter Systemprobleme und von Lernprozessen, die darauf antworten."

Das Zusammenspiel von Entwicklungslogik und Entwicklungsdynamik wird bei *Habermas* in folgenden Erklärungsschritten analysiert[66]:

1. *Ausgangszustand*: Das historisch gegebene Organisationsprinzip besitzt die notwendigen Kapazitäten, um Probleme der Systemintegration

und der Sozialintegration zu lösen. Beispiel: Politische Klassenstrukturen der mittelalterlichen Feudalgesellschaft waren den Kontrollproblemen feudaler Agrarproduktion und städtischen Handwerks durchaus gewachsen.

2. *Evolutionäre Herausforderungen:* Die Sozialstruktur produziert neuartige Systemprobleme, die mit der durch das hergebrachte Organisationsprinzip begrenzten Steuerungskapazität nicht bewältigt werden können. Beispiel: Wirtschaftliche Probleme (Fernhandel und Geldwirtschaft) überfordern die Anpassungs- und Lernfähigkeit der mittelalterlichen politischen Gesellschaft im Rahmen ihres Organisationsprinzips.

3. *Experimentierphase:* Kognitive Potentiale, die innerhalb der kulturellen Sphäre autonom nach ihrer eigenen Entwicklungslogik hervorgebracht worden sind, werden nun in einer experimentellen Weise für die soziale Organisation ausprobiert. Beispiel: Die Institutionalisierung universalistischer Strukturen innerhalb engumrissener Systemgrenzen für strategisches Handeln (Markt, rationale Organisation).

4. *Stabilisierung:* Im Erfolgsfall wird das neue Organisationsprinzip gesellschaftsweit institutionalisiert, besonders in grundlegende Rechtsstrukturen inkorporiert. Beispiel: Schaffung der Komplementärbeziehung zwischen kapitalistischer Wirtschaft, Privatrechtssystem, Steuerstaat und moderner Verwaltung.

Was diesen hier skizzierten Begriff des Organisationsprinzips für unsere Zwecke so fruchtbar macht, ist sein Bezug auf das Verhältnis von Rechtsstrukturen und Sozialstrukturen. Grundlegende Rechtsstrukturen erscheinen darin als die institutionelle Verkörperung gesellschaftlicher Organisationsprinzipien und können damit aus dem Zusammenspiel von kulturell-normativen und basisgesellschaftlichen Strukturen erklärt werden. Die Autonomie der Rechtsentwicklung bei *Nonet/Selznick* taucht hier als eigensinnige Entwicklungslogik normativer Strukturen wieder auf und läßt sich in *Habermas'* Modell "rational nachkonstruieren", was der Selbstreferenz im Rechtssystem gerecht wird. Zugleich aber wird sie um ein System/Umwelt-Modell ergänzt, das den Einfluß der gesellschaftlichen Entwicklungsdynamik auf normative Strukturen einer funktionalistischen Analyse (Systemprobleme Problemlösungskapazität) zugänglich macht. Die Kombination beider Modelle führt *Habermas* zu der für die Kovariation von rechtlichen und gesellschaftlichen Strukturen grundlegenden These, "daß in der sozialen Evolution höhere Integrationsniveaus nicht etabliert werden können, bevor sich nicht Rechtsinstitutionen herausgebildet haben, in denen ein moralisches Bewußtsein der konventionellen bzw. der postkonventionellen Stufe verkörpert ist"[67]. Damit wird das Problem, das bei *Nonet/Selznick* nur unbefriedigend angegangen wurde, wie es nämlich möglich ist, Rechtsentwicklung als autonome Entwicklung und zugleich in ihren gesellschaftlichen Bezügen darzustellen, von *Habermas* derart gelöst, daß er zwei unterschiedliche analytische Modelle einführt und anschließend miteinander kombiniert. *Habermas'* Vorgehen ermög-

licht es, den Begriff der Rechtsrationalität – wie oben schon angedeutet – in dreifacher Weise aufzufächern und damit in seinen innerrechtlichen wie in seinen gesellschaftlichen Bezügen darzustellen: Rationale Rekonstruktion erlaubt eine Analyse der jeweiligen "Normrationalität" des Rechts, welche die Möglichkeit der Rechtfertigung von Normen und Werten innerhalb einer gegebenen Rechts- und Moralordnung definiert. System/Umwelt-Analysen ergänzen diese innere Sicht um die "Systemrationalität", die die Problemlösungskapazitäten der Rechtsordnung gegenüber gesellschaftlichen Kontrollproblemen definiert. Aus den Anforderungen beider Aspekte – Normrationalität und Systemrationalität – an die inneren Strukturen des Rechtssystems bestimmt sich dann die "interne Rationalität" des Rechts, also seine spezifischen Grundbegrifflichkeiten und dogmatischen Systematisierungen.

Die Prozesse der "institutionellen Verkörperung" allerdings, die Frage, wie denn gesellschaftliche Organisationsprinzipien in Rechtsstrukturen "übersetzt" werden, bleibt bei *Habermas* ziemlich im Dunkeln. Hier mag der selektive Zugriff auf konkurriende Theorien weiterhelfen. Um die Umsetzung von Gesellschaftorganisation in konkretere Rechtsstrukturen analysieren zu können, kann man sich einiges von *Luhmanns* Konzept der "adäquaten Komplexität des Rechtssystems" erhoffen.

*Luhmann* weist zentrale Annahmen des klassischen Evolutionismus zurück, wie Unilinearität, Notwendigkeit, Fortschritt, und entwickelt eine minimalistische Version einer Evolutionstheorie, die auf drei Grundannahmen, nämlich über Dynamik, Mechanismen und Gerichtetheit zurückgeführt werden kann[68]. In letzter Instanz gewinnen danach evolutionäre Prozesse ihre Dynamik aus der fundamentalen Komplexitätsdifferenz zwischen Systemen und Umwelt[69]. Die Überproduktion an Möglichkeiten in der Umwelt resultiert in Anpassungsprozessen sozialer Systeme, wenn und soweit sie selbst stabile Evolutionsmechanismen für Variation, Selektion und Stabilisierung entwickeln[70]. Im Falle des Rechtssystems übernehmen normative Strukturen die Variation, institutionelle Strukturen (besonders Verfahren) die Selektion und dogmatisch-begriffliche Strukturen die Stabilisierung[71]. Gesellschaftlich-rechtliche Evolution ist dann gekennzeichnet durch die Interaktion zwischen diesen "endogenen" Evolutionsmechanismen einerseits und "exogener" Evolution der gesellschaftlichen Umwelt andererseits[72]. Und zwar wird die endogene Evolution des Rechts von exogenen Entwicklungen in der Weise systematisch beeinflußt, daß bestimmte Organisationsprinzipien der Gesamtgesellschaft das relative Gewicht der endogenen Evolutionsmechanismen (Normen, Verfahren, Dogmatiken) unterschiedlich verstärken bzw. abschwächen und damit die innere Dynamik im Rechtssystem extern beeinflussen. Das Recht wird an verschiedene Entwicklungsstadien gesellschaftlicher Differenzierung angepaßt und zwar in der Weise, daß das dominante Organisationsprinzip der Gesellschaft (Segmentierung, Stratifizierung, funktionale Differenzie-

rung) typische Konfigurationen im Rechtssystem schafft, für die sich jeweils spezifische Entwicklungsengpässe (mit dem entsprechenden "Krisen" – Potential) ausmachen lassen[73].

In segmentierten Gesellschaften, die sich durch Alternativenarmut auszeichnen, bestehen die evolutionären Probleme des "archaischen Rechts" darin, ausreichende Varietät an normativen Strukturen zu produzieren. Dieses Varietätsproblem wird erst in stratifizierten Gesellschaften mit einer differenzierten hierarchischen Ordnung gelöst. Sie sind in der Lage, einen größeren Normenreichtum zu schaffen, jedoch ist das "Recht der hochkultivierten vorneuzeitlichen Gesellschaften" Problemen der Selektion ausgesetzt. Funktional differenzierte Gesellschaften schließlich sind durch eine massive Überproduktion von Normen gekennzeichnet; das "positive" Recht hat zwar seinerseits das Selektionsproblem mit der Institutionalisierung von hochentwickelten Rechtsverfahren bewältigt, es verfügt jedoch in den herkömmlichen dogmatisch-begrifflichen Strukturen nur über unzureichende Stabilisierungsmechanismen. Die Krise des positiven Rechts liegt für *Luhmann* genau in jener Insuffizienz der dogmatischen Begrifflichkeit, die noch primär auf Rechtsanwendung zugeschnitten ist, begründet[74]. Ihr rigider Normativismus hindert das Entstehen eines "lernenden Rechts". Was nach *Luhmann* fehlt, ist eine "eigentlich rechtspolitische Begrifflichkeit, die es gestatten würde, verschiedene Problemlösungen in ihren Konsequenzen zu vergleichen, kritische Erfahrungen zu machen, Erfahrungen aus verschiedenen Rechtsbereichen zu vergleichen, kurz: zu lernen"[75].

Was kann die Theorie des "responsive law" von diesen beiden skizzierten Konzepten – Organisationsprinzip und adäquate Komplexität – gewinnen? In einer Art Gedankenexperiment wollen wir *Nonet/Selznicks* Konzept der inneren Entwicklungsdynamik des Rechts mit beiden konfrontieren, um zu sehen, wie sich jedes der drei Stadien der Rechtsentwicklung spezifisch verändert, wenn man die externe gesellschaftliche Dimension hinzunimmt. Angesichts unseres besonderen Interesses an moderner Rechtsrationalität sollen dabei die Modifikationen von repressivem und autonomem Recht nur kurz skizziert werden, während die Implikationen für responsives Recht genauer auszuarbeiten sind.

## 3. Neuinterpretation von repressivem und autonomem Recht

Bei *Nonet/Selznick* beginnt die Rechtsentwicklung mit "repressivem Recht" – einem Rechtstyp, dessen Primärfunktion sich auf die Legitimation von politischer Herrschaftsordnung beschränkt[76]. Aus der Sicht gesamtgesellschaftlicher Evolution müßte hier schon eine erste Korrektur ansetzen. Denn sowohl in *Habermas'* als auch in *Luhmanns* Theorie würde "repressives Recht" ein ziemlich modernes Entwicklungsstadium darstellen, welches das Organisationsprinzip einer "politischen Gesellschaft"

reflektiert[77]. Das legt es nahe, *Nonet/Selznicks* Typologie sozusagen nach rückwärts um einen weiteren, einen vor-modernen Rechtstyp – "archaisches Recht"[78] – zu erweitern, dessen Besonderheit in keinem der drei Rechtstypen von *Nonet/Selznick* aufgeht. Im Unterschied zu repressivem Recht verkörpert archaisches Recht das Organisationsprinzip segmentierter Gesellschaften, die durch das Vorherrschen von Familien- und Stammesstrukturen gekennzeichnet sind[79]. Vergeltung und Reziprozität sind die wesentlichen Prinzipien des archaischen Rechts, das in seinen zahlreichen Varianten von heiligem Recht sehr konkrete und rigide Normen aufweist, nur über ritualistische Formen des Rechtsverfahrens verfügt und den Schwerpunkt auf expressive, nicht auf instrumentale Funktionen des Rechts legt[80]. Ein solches primitives Recht transformiert sich erst dann langsam in Formen eines "repressiven" Rechts, wenn in der Gesellschaft Systemprobleme auftauchen, die die Kontrollkapazität des Verwandtschafts-Prinzips übersteigen und einen evolutionären Druck in Richtung politische Organisation der Gesellschaft über Herrschaftsrollen ausüben[81].

Erst mit dem Auftauchen eines neuen gesellschaftlichen Organisationsprinzips (*Luhmann:* "Stratifizierung"[82], *Habermas:* "politische Klassenherrschaft"[83])entsteht also ein Bedarf dafür, daß primitives Recht sich zum "repressiven Recht" wandelt. *Nonet/Selznick* analysieren im Detail dessen intime Verbindung zur politischen Macht[84]. Dies entspricht durchaus *Luhmanns* Begriff vom "Recht vorneuzeitlicher Hochkulturen", dessen Strukturen den Primat der politischen Ordnung in stratifizierten Gesellschaften widerspiegeln und ihren hierarchischen Herrschaftsformen entsprechen[85]. *Luhmanns* Analysen fügen dem aber noch das Moment adäquater Komplexität hinzu, die dann erreicht wird, wenn Gerichtsverfahren institutionalisiert werden, die dem höheren Konfliktpotential in stratifizierten Gesellschaften gerecht werden können[86]. *Habermas'* Analyse des gesellschaftlichen Organisationsprinzips in stratifizierten Gesellschaften wiederum erhellt besonders die spezifische Form von Sozialintegration, die unter dem Vorzeichen von "politischer Klassenherrschaft" möglich wird. Wieder ist es der wichtige Entwicklungsschritt zur Einrichtung von Gerichtsverfahren, der das Ablösen einer präkonventionellen durch konventionelle Moralität ermöglicht[87].

"Das ist der Fall, wenn der Richter, statt als bloßer Schiedsrichter an die kontingenten Machtkonstellationen der beteiligten Parteien gebunden zu sein, nach intersubjektiv anerkannten, durch Tradition geheiligten Rechtsnormen urteilen kann; wenn er neben den konkreten Handlungsfolgen auch die Intention des Täters berücksichtigt; wenn er sich nicht länger von der Idee der Vergeltung für entstandenen Schaden und der Wiederherstellung eines status quo ante leiten läßt, sondern die Regelverletzung eines Schuldigen bestraft."

Die Krise des repressiven Rechts wiederum sollte in terms von rechtlich-gesellschaftlichen Interrelationen analysiert werden. Erst das Zusammenspiel von zwei Krisentendenzen, einer inneren Legitimationskrise des repressiven Rechts *(Nonet/Selznick)*[88] und einer äußeren Systemkrise (*Ha-*

*bermas*: wirtschaftliche Systemprobleme[89]; *Luhmann*: funktionale Differenzierung[90]) macht den Entwicklungsschub von repressivem zu autonomem Recht einsichtig.

"Autonomes Recht" erfüllt in *Nonet/Selznicks* Definition die Bedingung formaler Rationalität in *Max Webers* Sinne: Trennung von Recht und Politik, juristische Professionalisierung, strikte Regelorientierung, Universalismus und Präzision, artifizielle Argumentation und ein Begriff von Verfahrengerechtigkeit[91]. Aber erst *Habermas'* Analyse der Beziehungen zum gesellschaftlichen Organisationsprinzip würde diese "interne Rationalität" des formalen Rechts aus den Dimensionen von "Normrationalität" und Systemrationalität" erklären können[92]. Auf der einen Seite trägt das moderne Formalrecht zur Mobilisierung und Allokation von Ressourcen bei; diese Systemrationalität wird durch die Elemente der Konventionalität, des Legalismus und der Formalität gestützt. Auf der anderen Seite ventionelle Moralität zu entwickeln, nämlich die Notwendigkeit, Normen durch Rekurs auf universale Prinzipien zu rechtfertigen. Ganz komplementär lassen sich *Luhmanns* Analysen des "positiven Rechts" verwenden, um zu zeigen, inwieweit das Recht adäquate Komplexität in einer funktional differenzierten Gesellschaft entwickelt. In dieser Sicht scheinen die Positivität des Rechts und die Trennung rechtsanwendender und gesetzgebender Verfahren die ausschlaggebenden Variablen zu sein[93]. Für Krisentendenzen innerhalb des modernen Formalrechts gilt wieder das oben Gesagte. Die aus der inneren Dynamik des Rechtssystems erwachsenden Bedürfnisse nach einem inhaltlich legitimierten responsiven Recht *(Nonet/Selznick)* sind in ihrer Interdependenz zu sehen sowohl zu Komplexitätsdefiziten der Sachstrukturen des Rechts *(Luhmann)* als auch zu Krisentendenzen des dominierenden Organisationsprinzips *(Habermas)*[94].

Eine solche versuchsweise Re-formulierung von "repressivem" und "autonomem" Recht im gesellschaftlichen Kontext mag plausibel gemacht haben, daß der hier gewählte kombinatorische (oder auch: eklektische, harmonistische?) Ansatz die Erklärungskraft der Theorie des "responsive law" steigern kann. Es bringt jedenfalls beträchtliche Modifikatinen der Theorie mit sich, wenn man interne Konfigurationen des Rechts mit gesellschaftlichen Organisationsprinzipien in Beziehung setzt und sie auf die Adäquatheit ihrer inneren Komplexität befragt. Wie gezeigt, zwingt ein solcher Ansatz dazu, die dreifache Typologie der Rechtsentwicklung um einen vierten Typ – archaisches Recht – zu erweitern und gibt Anlaß zu einer veränderten Einschätzung von repressivem und autonomem Recht sowohl in bezug auf ihre inneren Strukturen als auch in bezug auf ihre äußeren gesellschaftlichen Funktionen. Das gleiche gilt für die Analyse der den Rechtstypen immanenten Entwicklungskrisen. Jedoch ist die uns in diesem Zusammenhang am meisten interessierende Frange noch unbeantwortet: In welcher Hinsicht kann ein solcher kombinatorischer Ansatz unser Verständnis von responsivem Recht bereichern?

## III. "The Quest for Responsive Law": Sozialadäquate Komplexität des Rechtssystems?

Um kurz den bisherigen Gedankengang zu rekapitulieren: Unsere Ausgangsfrage war: Welche Lösungsperspektiven werden in der Rechtstheorie angeboten, die einen Ausweg aus der gegenwärtigen Krise formaler Rechtsrationalität bieten? *Nonet/Selznick* präsentieren in der Kombination einer sozialwissenschaftlichen Strategie mit einem evolutionären Ansatz das Konzept des responsiven Rechts, das – in unserer Interpretation – zwei Elemente enthält: materiale und reflexive Rationalität (Abschnitt I). Ihr Entwicklungsmodell erscheint jedoch deshalb unbefriedigend, weil es nur unzureichend die Interdependenzen von Rechtsentwicklung und gesellschaftlicher Evolution thematisiert. Ein umfassenderes Modell der gesellschaftlich-rechtlichen Ko-variation läßt sich in der Weise konstruieren, daß die Analyse der "inneren Dynamik des Rechts" *(Nonet/Selznick)* mit Vorstellungen "externer" gesellschaftlicher Einflüsse auf das Recht kombiniert wird, genauer mit den Konzepten von "gesellschaftlichen Organisationsprinzipien" *(Habermas)* und von "adäquater Komplexität" des Rechts *(Luhmann)*. In diesem Zusammenhang konnten die Entwicklungsstadien des repressiven und des autonomen Rechts mit entsprechenden Stadien in der gesamtgesellschaftlichen Entwicklung systematisch verbunden werden (Abschnitt II). Die daraus resultierende Frage lautet, wie "responsives" Recht im Lichte rechtlich-gesellschaftlicher Interdependenzen re-formuliert werden müßte. Auf welche gesellschaftlichen Organisationsprinzipien antwortet "responsives" Recht? Welche Elemente des responsiven Rechts weisen eine solche innere Komplexität auf, daß sie als adäquat gegenüber den Anforderungen der Gesellschaft bezeichnet werden können? Und schließlich: Wenn es stimmt, daß die innere Rechtsdynamik Entwicklungschancen sowohl für materiale wie für reflexive Rationalität des Rechts produziert, wie steht es um deren Realisierungsmöglichkeiten im größeren sozialen Kontext?

Um solchen Fragen eine halbwegs beantwortbare Fassung zu geben, wollen wir sowohl aus *Habermas'* als auch aus *Luhmanns* Theorie die jeweiligen Konzepte einer post-modernen Gesellschaft herausarbeiten und dann die Frage stellen, wie sich im Lichte dieser Konzepte Sozialstrukturen auf materiale beziehungsweise auf reflexive Rechtsrationalität auswirken. Auch für dieses Vorgehen gilt das eingangs Gesagte, daß *Luhmann* und *Habermas* zwar die Protagonisten einer fundamentalen makro-soziologischen Kontroverse sind und entsprechend höchst unterschiedliche Einschätzungen post-moderner Gesellschaften bieten[95]. Dennoch darf man mit gutem Grund, jedenfalls für die jetzigen Analysezwecke, ihre Theorien als komplementär und nicht als konkurrierend behandeln. Denn im Kontext der Normen- und Rechtsevolution fokussieren sie beide ein und dasselbe Fundamentalproblem: Ist normative Integration in einer Gesell-

schaft noch möglich, die sich durch innere Widersprüche, durch desintegrierende, ja disruptive Konflikte zwischen unterschiedlichen Rationalitäten hochspezialisierter Subsysteme auszeichnet[96]? Auch ihre jeweiligen Antworten lassen sich als einanderergänzend betrachten. Auf der einen Seite sind *Habermas'* Analysen der Krisentendenzen im Spätkapitalismus in eine allgemeine Systemtheorie übersetzbar, die diese Krisen als Rationalitätskonflikte verschiedener gesellschaftlicher Subsysteme – Wirtschaft, Politik, Kultur, Familie – interpretiert[97]. Auf der anderen Seite sind die integrativen Mechanismen, auf die *Luhmann* setzt, in *Habermas* Theoriengebäude durchaus eingeschlossen. In *Habermas'* Terminologie handelt es sich dabei um Mechanismen der Systemintegration, die allerdings ohne ergänzende Mechanismen der Sozialintegration erfolglos bleiben müßten[98]. Dieses wechselseitige Sich-Einverleiben von Systemtheorie und Kritischer Theorie sollte uns gegenüber den Ausschließlichkeitsansprüchen in den Großkontroversen skeptisch stimmen und dazu ermutigen, das analytische Potential beider zu benutzen, um die Realisierungschancen des responsiven Rechts besser einschätzen zu können.

## 1. Materiale Rationalität des Rechts in der Krise des interventionistischen Staates

*Habermas* hat eine Theorie der Legitimationsprobleme im organisierten Kapitalismus entwickelt,[99] die man systematisch auf den Begriff des responsiven Rechts beziehen kann. In ihrem Kern besagt die Theorie, daß Krisen im Kapitalismus nicht wirklich gelöst werden, sondern im Laufe ihrer "Bewältigung" sich nur von einem gesellschaftlichen Teilsystem in das andere verschieben. Primäre ökonomische Krisen können zwar teilweise vom interventionistischen Staat abgefangen werden, dies aber produziert nur wiederum Krisenphänomene im politischen System. Die hier entstehenden politischen Legitimationsprobleme werden auf das soziokulturelle System abgeschoben, dessen Politisierung eine kulturelle Motivationskrise auslöst, die allenfalls durch fundamentale Änderungen der normativen Strukturen überwunden werden kann[100]. In diesem Analyserahmen wird die Krise der formalen Rationalität des Rechts – die *Nonet/Selznick* ausschließlich aus inneren Problemen des "autonomen" Rechts erklären – in Beziehung zu einem gesamtgesellschaftlichen Phänomen gesetzt: den politischen Interventionen des modernen Wohlfahrtsstaats. Die frühere Systemrationalität des Formalrechts, die sich im Zusammenspiel von entwickelter Marktstruktur, formalem Privatrechtssystem, Steuerstaat und bürokratischer Administration zeigt, wird in dem Maße interminiert, als der Staat Verantwortung übernimmt für die Korrektur von Marktdefiziten, für globale wirtschaftspolitische Steuerung und für kompensatorische Sozialpolitik[101]. "Materialisierung des Formalrechts" repräsentiert die Parallelentwicklungen im Rechtssystem. Die ma-

teriale Rechtsrationalität ist für sozialstaatliche Zwecke instrumentalisierbar, besonders in den Merkmalen des Partikularismus, der Ergebnisorientierung, des instrumentellen Normverständnis und der zunehmenden "Verrechtlichung" vormals rechtsfreier, autonomer Sozialprozesse. Sozialstaatliche Verrechtlichung "zähmt" das wirtschaftliche System durch Konstitutionalisierung und soll mit sozialstaatlichen Verbürgungen dessen externe Effekte auffangen[102].

Nach allem scheint also materiale Rechtsrationalität die sozialadäquate Rechtsform des Sozialstaates zu liefern. *Habermas'* Pointe jedoch ist das genaue Gegenteil, nämlich die Grenzen ihres Wachstums zu demonstrieren. *Habermas* identifiziert drei Krisentendenzen innerhalb des politischen Systems, die das Potential einer politisch-rechtlichen materialen Rationalität drastisch begrenzen[103]. Widersprüchliche Imperative eines ökonomischen Krisenmanagements sind geeignet, die Kontrollkapazitäten staatlicher Intervention schnell zu überfordern, sie stürzen das politische System in eine "Rationalitätskrise", welche die Systemintegration gefährdet und damit auch die Sozialintegration in Frage stellt[104]. Die Komplexität sozioökonomischer Presse läßt sich nach allen Erfahrungen nicht in der Weise von politisch-rechtlichen Kontrollmechanismen erfassen, wie es eine materiale Rationalität erfordern würde. Wichtiger für *Habermas* ist eine zweite Krisentendenz, die "Legitimationskrise" im organisierten Kapitalismus. Aufgrund ökonomischer Konzentrationsprozesse und staatsinterventionistischer Politiken verliert der Markt seine einst selbstverständliche Legitimation als ein Verteilungsmechanismus, der sozusagen "natürliche" und damit auch gerechte Ergebnisse produzierte. Nun wird das Ausmaß staatlicher Markt-Subsitution und Markt-Kompensation eine Frage bewußter politischer Entscheidung, die dann aber das politische System in verstärktem Maße legitimationsbedürftig und gerade in der nur schwer beherrschbaren politischen Steuerung der Wirtschaft von Massenloyalität abhängig macht. Der naheliegende Ausweg, legitimierende Ersatz-Ideologien künstlich zu produzieren, steht nach *Habermas* dem politischen System deshalb nicht offen, weil resistente normative Strukturen im kulturellen System einer manipulativen Ideologisierung unüberwindliche Hindernisse bereiten[105]. Die Konfrontation politischer Legitimationsbedürfnisse mit der normativen Entwicklungslogik im sozio-kulturellen System produziert also mit Notwendigkeit eine "Motivationskrise", die der materialen politischen und rechtlichen Rationalität des Staatsinterventionismus eine effektive Grenze setzt[106].

Erst eine neuartige "diskursive" Rationalität, die sich nach *Habermas* in autonomen Entwicklungsprozessen der normativen Sphäre herausbildet, könnte, sofern ihre gesellschaftsweite Institutionalisierung gelingen sollte, die Legitimationsprobleme des modernen Staates lösen[107]. Zugrunde liegt dem eine Theorie politischer Legitimation, die aufgrund irreversibler normativer Entwicklungen ausschließlich prozedurale Legitimations-

modi für möglich hält: "Jetzt, da letzte Gründe theoretisch nicht mehr plausibel gemacht werden können, *erhalten die formalen Bedingungen der Rechtfertigung selber legitimierende Kraft.* Die Prozeduren und Voraussetzungen vernünftiger Einigung werden selbst zum Prinzip"[108]. Demgegenüber hängt die konkrete Frage der Institutionalisierung, also die Frage, welche Organisationsstrukturen und welche Diskussions- und Entscheidungsmechanismen prozedural legitime Ergebnisse produzieren können, vom sozialen Kontext ab, "von der konkreten gesellschaftlichen Ausgangssituation, von gegebenen Interessenlagen, von Dispositionsspielräumen, Informationen usw."[109] *Habermas'* eigene Vorschläge zur Institutionalisierung prozeduraler Legitimation schließen insbesondere Mechanismen organisationsinterner Demokratie in öffentlichen Verbänden, Gewerkschaften und Funktionseliten ein, partizipatorische Mechanismen in verschiedenen gesellschaftlichen Subsystemen, hauptsächlich im Erziehungs- und Kulturbereich, und das "pragmatische Modell" eines Dialogs für die institutionalisierte Zusammenarbeit im Dreieck von Wissenschaft, Politik und Öffentlichkeit[110].

Dieses Programm einer Demokratisierung gesellschaftlicher Subsysteme könnte seine rechtlichen Entsprechungen im Konzept des "responsive law" finden. Die Re-politisierung des Rechts, die Ausdehnung demokratischer Partizipation in den Rechtsprozeß hinein und der "institutional design" für die Organisation unterrepräsentierter Interessen bei *Nonet/Selznick*, wären rechtliche Gegenstücke gesellschaftlicher Demokratisierung. Die Parallele stimmt aber eben nur für solche Elemente im responsiven Recht, die wir "reflexiv" genannt haben. Diese unterschiedliche Einschätzung materialen und reflexiven Rechts wird bei *Habermas* in seiner Kritik an *Max Weber* besonders deutlich. *Habermas* wendet sich gegen *Webers* undifferenzierte Verwendung des Begriffs "antiformaler Tendenzen", mit der sowohl partikularistische Re-ideologisierungen des Rechts als auch die weitergehende Verkörperung posttraditionaler Bewußtseinsstrukturen im Recht unterschiedslos belegt und abqualifiziert werden[111]. In *Habermas'* Einschätzung handelt es sich dabei um zwei fundamental verschiedene Wandlungen der Rechtsstruktur, die unserer Unterscheidung materialer und reflexiver Tendenzen nahekommt. *Habermas'* Analyse der Krisentendenzen im Interventionsstaat legt damit folgende Schlußfolgerung nahe. Konfrontiert man das Konzept des "responsive law" mit den von *Habermas* herausgearbeiteten Krisenphänomenen, so sind die Entfaltungschancen der unterschiedlichen Tendenzen dieses Rechtstyps durchaus auch unterschiedlich zu beurteilen. Tendenzen in Richtung einer materialen Rechtsrationalität laufen recht bald an den strukturellen Grenzen des Staatsinterventionismus auf. Die Grenzen, sowohl der Kontrollkapazität staatlicher Steuerung als auch von deren öffentlicher Legitimation kann eine Materialisierung des Rechts nicht aufbrechen.

In seinen Tendenzen zu einer eher reflexiven Rationalität hingegen

kann "responsive law" Möglichkeiten einer institutionalisierten prozeduralen Legitimation entwickeln, die – in *Habermas'* Einschätzung – das dominante Organisationsprinzip einer post-modernen Gesellschaft repräsentieren könnte. Recht würde dann nicht als sozialtechnisches "Steuerungsmedium" fungieren, das sich an den eigensinnigen Strukturen der Lebenswelt bricht oder aber diese Strukturen zerbricht, sondern als legitimationsbedürftige und -fähige "Institution", die sich darauf beschränkt, die "äußere Verfassung" der Bereiche von Sozialisation, sozialer Integration und kultureller Reproduktion zu normieren, und die diese in ihren selbstregulatorischen Lern- und Kommunikationsprozessen nicht gefährdet[112].

## 2. Reflexive Rationalität und funktionale gesellschaftliche Differenzierung

Bis hierhin haben wir die Denkmodelle der Kritischen Theorie dazu benutzt, die Realisierungschancen des responsive law im gesellschaftlichen Kontext einzuschätzen. Wenn wir nun die gleiche Frage in die Sprache der neo-funktionalistischen Systemtheorie übersetzen, in der *Luhmann* die *Durkheim-Parsons*-Tradition weiterverfolgt, so werden wir auf ganz ähnliche Problemkonstellationen stoßen – wenn auch in noch abstrakterer und umfassenderer Perspektive. Krisentendenzen im organisierten Kapitalismus erscheinen in diesem Lichte nur als Sonderfälle eines allgemeineren Phänomens. Nicht mehr nur die "evolutionären Herausforderungen ungelöster, ökonomisch bedingter Systemprobleme", wie *Habermas* es sieht, sondern allgemeiner die gesamtgesellschaftliche Umstellung auf funktionale Differenzierung bestimmt nun die Entwicklungsdynamik, die hochspezialisierte gesellschaftliche Subsysteme dazu treibt, ihre Eigenrationalität so extrem zu forcieren, daß krisenhafte Inter-System-Konflikte unvermeidlich sind[113]. Die bei *Habermas* so zentrale "Motivationskrise", die aus dem Grundwiderspruch zwischen der Logik des politisch-ökonomischen Systems und der Logik normativer Strukturen entsteht, erscheint nun als ein eher marginales Problem, sofern man eine Reihe von fundamentaleren Konflikten ins Auge faßt: "etwa die zwischen dem bereits welteinheitlich konstituierten Gesellschaftssystem und den noch auf regionaler Integrationsbasis festgehaltenen politischen Teilsystemen; oder die Folgeprobleme der Ersetzung religiöser durch monetäre Zukunftsvorsorge; oder die Folgen des Abbaus von Schichtung; oder die zunehmenden Wertungs- und Planungsdivergenzen zwischen wissenschaftlicher Planung und ökonomischer Produktionssteuerung; oder die mit strukturellen Interdependenzen zunehmenden Tempoanforderungen, denen Sozialisations-, Ausbildungs- und Institutionalisierungsprozesse nicht zu folgen vermögen"[114]. In dieser Perspektive gewinnt man auch schließlich eine andere Sicht auf die Konturen eines neuen "responsiven" Rechts. Das Grundproblem lautet nun-

mehr: Wie partizipiert und wie reagiert Recht in den säkularen Prozessen funktionaler Differenzierung[115]?

Der Ausdruck "formale" Rationalität im Sinne *Max Webers* erscheint in dieser Sicht ziemlich irreführend, da Form und Substanz in einem systemtheoretisch orientierten Vergleich traditionellen und modernen Rechts fast austauschbar sind[116]. Eher schon angemessen scheint "autonomous law", da es auf die entscheidende funktionale Ausdifferenzierung des Rechtssystems verweist: Steigerung der Autonomie des Rechts, seine Trennung von Moralität und Wahrheit, seine relative Unabhängigkeit von Macht und Politik. Mit Hilfe dieser Autonomisierungstendenzen lassen sich die Merkmale des "Formalismus" im Recht erklären: die strikte Regel-Orientierung, professionalisierte "künstliche" Argumentationsweisen, die Prominenz von Verfahren. Und die Krise des "autonomous law" hat in genau diesen Prozessen funktionaler Differenzierung ihre Ursache. Das Rechtssystem hat sich, besonders in seinen begrifflich-dogmatischen Strukturen, noch nicht den Anforderungen einer hochdifferenzierten Gesellschaft gestellt. Rechtsdogmatik ist immer noch dem klassischen Rechtsanwendungsmodell in richterlicher Perspektive verhaftet und hat noch keinen Begriffsapparat entwickelt, der den Planungs- und Politikerfordernissen in den Beziehungen zwischen spezialisierten Teilsystemen angemessen wäre[117].

"Materiale" Rationalität in *Max Webers* Sinne, die der Krise des Formalrechts mit einer Re-moralisierung und einer Re-politisierung des Rechts begegnen will, erscheint in dieser Perspektive klar als regressive Tendenz. Eine neuerliche Verschmelzung des Rechts mit Wissenschaft, Moral, Politik würde die spezifisch juridische Rationalität zerstören, ohne daß sie sie durch eine neue erfolgversprechende Orientierung ersetzte[118]. In *Luhmanns* Sicht – die in diesem Aspekt der von *Habermas* weitgehend entspricht – würde eine durchgehende Re-materialisierung des Formalrechts das politisch-rechtliche System unvermeidlich in eine Steuerungskrise stürzen. Denn in den Prozessen funktionaler Differenzierung haben die sozialen Teilsysteme einen solchen Grad an innerer Komplexität gewonnen, daß keines der gesellschaftlichen Teilsysteme – sei es Politik, Wissenschaft, Wirtschaft, Moral, Recht, sei es eine Kombination zwischen ihnen – mehr in der Lage ist, die notwendige Kontrollkapazität zu ihrer inneren Steuerung zu entwickeln. Responsives Recht in seiner Dimension materialer Rechtsrationalität würde danach, sofern es gesellschaftsweit wirksam werden sollte, eher zu einer regressiven Entdifferenzierung gesellschaftlicher Strukturen führen als zu ihrer normativen Integration.

Das führt zu der zentralen Frage wie denn eine Integration der Gesellschaft noch möglich ist unter Bedingungen extremer funktionaler Differenzierung? Und welche Rolle Recht denn spielt als integrativer Mechanismus[119]? *Luhmanns* Antwort unterscheidet sich beträchtlich von

klassischen Lösungen, etwa *Durkheims* bekannter "organischer Solidarität", die durch "restitutives Recht" gestützt wird[120]. Für *Luhmann* stellt organische Solidarität immer noch einen traditionellen Mechanismus dar, der noch nicht wirklich radikale Folgerungen aus gesellschaftlicher Arbeitsteilung gezogen hat, da diese Solidarität immer noch auf zwar hochgeneralisierte, aber doch für alle Subsysteme gemeinsame gesellschaftliche Normen setzt. Für ausdifferenzierte gesellschaftliche Subsysteme, die in äußeren Funktionen und internen Strukturen extrem differieren, kann Integration nicht mehr durch politisch-rechtliche Definition uniformer normativer Strukturen geleistet werden. *Luhmann:* "Die Gesellschaft kann bei zunehmender Komplexität immer weniger garantieren, daß alle Teilsysteme unter gleichen Strukturen gleichförmig operieren. . . Integration muß vielmehr dadurch vermittelt werden, daß *alle Teilsysteme füreinander innergesellschaftliche Umwelt sind*"[121]. Daraus folgt, ". . .daß alle Teilsysteme der Gesellschaft. . . nicht nur ihre eigene Funktion adäquat erfüllen müssen, sondern auch als gesellschaftliche Umwelt anderer Systeme zu deren Funktionen und strukturellen Errungenschaften in einem sinnvollen Verhältnis der Kompatibilität stehen müssen"[122]. Damit verlangt funktionale Differenzierung eine Verlagerung integrativer Mechanismen von der Ebene der Gesamtgesellschaft auf die Ebene der Subsysteme[123]. Eine zentralisierte Integration über politisch-rechtliche Mechanismen – wie sie in stratifizierten Gesellschaften noch möglich war – ist heute effektiv ausgeschlossen und kann nicht durch andere, etwa rechtliche, moralische oder wissenschaftliche Integrationsmittel ersetzt werden. Wenn unter modernen Bedingungen gesellschaftliche Integration bedeutet zu vermeiden, daß die Maximierung einer Subsystem-Rationalität zu unlösbaren Problemen für andere Funktionssysteme führt, dann ist eine dezentrale Integration unvermeidlich. "*Entsprechende Beschränkungen müssen in die Reflexionsstruktur eines jeden Funktionssystems eingebaut sein, soweit sie sich nicht unmittelbar aus den laufenden Beziehungen zur Umwelt ergeben*"[124]. Sonach scheinen "reflexive" Strukturen der Schlüssel zu unserem Problem zu sein, die Rolle des "responsive law" in funktional differenzierten Gesellschaften zu bestimmen. Dies bedarf einiger Vertiefung.

Reflexionsprozesse innerhalb gesellschaftlicher Teilsysteme müssen in hochdifferenzierten Gesellschaften dann an die Stelle gesamtgesellschaftlicher Integrationsmechanismen treten, wenn die Teilsysteme widersprüchliche Orientierungen herausbilden, die nur noch systemintern harmonisiert werden könne.[125]. Nach *Luhmann* kann jedes Teilsystem seine selektiven Operationen an drei Beziehungen ausrichten: 1) als Funktion in Richtung auf die Gesamtgesellschaft, 2) als Leistung in bezug auf andere gesellschaftliche Teilsysteme, 3) als Reflexion in bezug auf sich selbst[126]. Entscheidend ist nun, daß diese Orientierungen einander widersprechen und nicht unter einem gemeinsamen Oberzweck subsumiert werden können. Im politischen System etwa besteht ein inneres Spannungsverhältnis

zwischen der gesellschaftlichen Funktion der Politik (Herstellung und Durchsetzung gesellschaftlich bindender Entscheidungen) und ihren Leistungen an andere Sozialsysteme (Bereitstellung von Machtressourcen und ausreichende Legitimation der Entscheidungen). Dieses Spannungsverhältnis kann nur noch intern durch politische Reflexionsprozesse (historische Identifizierung) in ein Gleichgewicht gebracht werden[127]. Ganz analog ist es die Aufgabe von Reflexionsstrukturen in jedem anderen gesellschaftlichen Teilsystem, den Konflikt zwischen Funktion und Leistung dadurch zu entschärfen, daß den Kapazitäten des Systems innere Schranken auferlegt werden, und zwar im Interesse ihrer Eignung als Komponenten der Umwelt anderer Teilsysteme[128]. "Reflexion muß dann zwischen Funktion und Leistung vermitteln, weil die Gesellschaft für das Funktionssystem sowohl übergreifendes System als auch soziale Umwelt ist"[129]. Aufgrund ihrer Ausgleichfunktion werden also Reflexionsstrukturen innerhalb der Teilsysteme zu Hauptträgern gesellschaftlicher Integration. Auf unser Problem des "responsive law" zurückgewendet bedeutet dies, daß aus dieser Perspektive das Recht reflexive Dimensionen entwikeln muß, wenn es als integrativer Mechanismus überhaupt noch in Frage kommen soll[130].

### 3. Der Konvergenzpunkt: Reflexionsstrukturen und demokratischer Diskurs in gesellschaftlichen Teilsystemen

Gerade in der reflexiven Orientierung des "responsive law" vermuten wir den Konvergenzpunkt der drei diskutierten Entwicklungsmodelle des Rechts. Unsere Übersetzungen von "responsive law" in die Sprache der Kritischen Theorie und des Neo-funktionalismus haben zunächst Gemeinsamkeiten in der skeptischen Einschätzung von rechtlichen Materialisierungstendenzen erbracht. Darüber hinaus aber lassen sich in unserer Interpretation des responsiven Rechts systematische Verbindungen zwischen prozeduralem Legitimationstyp bei *Habermas* und systeminternen Reflexionsstrukturen bei *Luhmann* herstellen. Unserer These ist, *daß einerseits Reflexion in gesellschaftlichen Teilsystemen erst gelingen kann, wenn Demokratisierungsprozesse in diesen Teilsystemen diskursive Strukturen herstellen, und daß andererseits die primäre Funktion von Demokratisierung gesellschaftlicher Teilbereiche nicht in Partizipationssteigerung oder Machtneutralisierrung liegt, sondern in der systeminternen Reflexion auf gesellschaftliche Identität.* Reflexive Rechtsstrukturen vermögen in diesem Kontext eine nicht unwesentliche Rolle zu spielen. Wir stoßen hier auf komplementäre Begriffe von relexiver Rationalität des Rechts: Während die neo-funktionalistische Theorie auf eine neuartige Selbst-Beschränkung des Rechts hinführt, verweist die Kritische Theorie auf ein wichtiges Potential reflexiven Rechts, interne demokratische Prozesse in sozialen Subsystemen über Verfahren und Organisation zu fördern.

Was eine solche *rechtliche Selbst-Beschränkung* bedeuten kann, wird klarer, wenn man die oben erwähnte Typologie möglicher Orientierungen – Funktion, Leistung, Reflexion – auf das Rechtssystem anwendet. Als Funktion des Rechts kann formuliert werden, daß es für die Gesellschaft normative Strukturen in Form von kongruent generalisierbaren Erwartungen bereitstellt[131]. Die Leistungen des Rechtssystems hingegen bestehen darin, Konflikte zu regulieren, die in anderen sozialen Systemen als nicht mehr mit systemeigenen Mitteln lösbar erscheinen. Beide Orientierungen überschneiden sich natürlich, aber sie können einander deutlich widersprechen. Die Produktion kongruenter Generalisierung mag nicht ausreichen, um Regeln für konkrete Konfliktlösungen bereitzustellen. Umgekehrt mag das Rechtssystem im Laufe seiner Konfliktlösungen Regeln entwickeln, die sich als nicht kongruent generalisierbar erweisen. Die Rolle rechtlicher Reflexion muß dann darin bestehen, das Spannungsverhältnis zwischen Rechtsfunktion und Rechtsleistung zu entschärfen, indem sie den inneren Kapazitäten des Rechtssystems Beschränkungen auferlegt.

Unter modernen Bedingungen – so kann man vermuten – wird das im wesentlichen bedeuten, daß das Recht seine Leistung in bezug auf andere Sozialsysteme einschränken muß. Statt umfassender Konfliktregulierung in allen möglichen Sozialkontexten im Sinne materialer Rechtsrationalität, die letztlich eine Selbstüberforderung des Rechts bedeuten, müßte Reflexion im Rechtssystem zu beschränkteren, abstrakteren, indirekteren Formen der sozialen Kontrolle führen.

Der entscheidende Punkt liegt in der *strukturellen Entsprechung von Rechtsnormierungen einerseits und den sozialen Regelungssituationen in gesellschaftlichen Subsystemen andererseits.* In den bekannten Materialisierungsprozessen wurde auf dieses Verhältnis von "opportunity structure" in der Gesellschaft und der "conceptual readiness" des Rechts wenig Rücksicht genommen. Man tendierte dazu, Regulierungstechniken auch dort zu benutzen, wo die Sozialstrukturen sich solchen Regulierungen kaum fügen. Aspekte des Erziehungssystems in der Sozialpolitik sind dafür schlagende Beispiele. Die aktuelle Kritik an Verrechtlichungstendenzen hat immer wieder aufs neue herausgearbeitet, daß materiale Rechtsprogramme als sozialstaatliche Steruerungsmedien Funktionsmodi, Rationalitätskriterien und Organisationsformen besitzen, die der lebensweltlichen Struktur der regulierten Sozialbereiche nicht angemessen sind, an ihr erfolglos auflaufen oder aber um den Preis ihres Erfolges diese zerstören[132]. In reflexiver Orientierung hingegen würde man die Regulierungskapazitäten des Rechts selbstkritisch prüfen, ehe man die Dynamik der "Verrechtlichung" einleitet. Leitfrage wäre dann nicht länger, ob Sozialkonflikte bestehen, demgegenüber ein sozialkompensatorisches Recht "responsiv" sein müßte. Vielmehr, ob in Sozialkonflikten eine "opportunity structure" besteht, die Rechtsinterventionen erfolgversprechend erscheinen lassen und zwar erfolgversprechend in dem Sinne, daß sie nicht um den Preis

effektiver Intervention gewachsene Sozialstrukturen irreversibel zerstören.

Worauf könnte nun eine solche Selbstlimitierung des Rechts hinauslaufen? Lassen sich plausible Aussagen darüber machen, in welche Richtung autonome Reflexion im Recht die Rolle des Rechtssystems gegenüber anderen Teilsystemen definiert?

Resultat solcher rechtlichen Reflexionsprozesse wäre nun nicht notwendig eine Politik der "de-regulation", eine Re-formalisierung des materialen Rechts. Der rechtliche "self-restraint" ging womöglich in ganz andere Richtung, nämlich das Recht darauf zu konzentrieren, strukturelle Voraussetzungen für selbstregulatorische Prozesse in anderen Sozialzusammenhängen zu schaffen. Dies bedeutet nicht nur rechtliche Garantien für Systemautonomie anderer Sozialsysteme. Solche Garantien sind zwar wichtig, erschöpfen aber nicht das Potential eines reflexiven Rechts. Es ist eher *Habermas'* Konzept einer *Demokratisierung gesellschaftlicher Teilsysteme*, das mit seiner Betonung prozeduraler Legitimation die Richtung angeben kann, auf die reflexives Recht sich hinentwickeln mag. In seiner Kritik an Verrechtlichungsprozessen im Sozialstaat hat *Habermas* die schon oben erwähnte wichtige Unterscheidung zwischen Recht als "Medium" und Recht als "Institution" getroffen. Während Recht als sozialstaatliches "Medium" der Gesellschaftssteuerung die kommunikativen Strukturen des verrechtlichten Handlungsbereichs zu verletzen droht, kann Recht als "Institution" kommunikative Strukturen unverletzt lassen, ja, sie sogar rechtlich fördern, wenn es sich auf die – wie *Habermas* sagt – äußere Verfassung" des kommunikativ strukturierten Sozialbereichs beschränkt. Recht als "äußere Verfassung" muß Verfahren der Konfliktregulierung ermöglichen, die den Strukturen verständigungsorientierten Handelns angemessen sind: "diskursive Willensbildungsprozesse und konsensorientierte Verhandlungs- und Entscheidungsverfahren"[133]. Genau dies trifft unser Verständnis der gesellschaftlichen Leistungen eines "reflexiven" Rechts: die äußere Verfassung für Prozesse der Selbstreflexion in anderen Sozialbereichen zu normieren. Denn zur Herstellung prozeduraler Legitimität, die in unterschiedlichen Sozialkontexten auch unterschiedliche Organisationsformen von "Diskursen" benötigt, kann Recht nicht unwesentliche Beiträge leisten. Das Rechtssystem kann Normierungen für Verfassungen, Verfahren, Organisation und Kompetenzen entwickeln, die andere Sozialsysteme als Voraussetzung demokratischer Selbst-Organisation und Selbst-Regulierung benötigen. Anstatt, wie es unter materialer Rechtsrationalität tendenziell geschieht, die Funktionen anderer Sozialsysteme autoritativ zu definieren, beziehungsweise deren Input- und Output-Leistungen rechtlich zu regulieren, müßte das Recht nun seine Aufmerksamkeit auf solche Mechanismen richten, *die systematisch ihrerseits Reflexionsstrukturen innerhalb anderer Subsysteme fördern.*

Der wesentliche Punkt in diesem Zusammenhang ist die subsystemspe-

zifische Fassung der allgemeinen Diskurs-Theorie. Ebenso wie es unmöglich ist, durch zentralisierte politische Entscheidungsprozesse eine funktional differenzierte Gesellschaft normativ zu integrieren, ist es unmöglich, für subsystemspezifische Reflexionsprozesse eine universelle Rechtfertigungsstruktur, eine generelle Diskursmoral, ein allgemeingültiges Entscheidungsverfahren zu entwickeln. Reflexionprozesse im Wirtschaftssystem bedürfen prinzipiell anderer "subsystemadäquater" Normierungen von Verfahren und Organisation als etwa im Erziehungssystem – und enstsprechend bedarf es anderer subsystemadäquater Leistungen eines reflexiven Rechts[134]. Recht müßte sich dann darauf beschränken, demokratische selbstregulatorische Mechanismen zu installieren, zu korrigieren und zu redefinieren, und würde gerade nicht die Gesamtverantwortung für soziale Ergebnisse übernehmen. Die technischen Mittel, die das Rechtssystem hierzu einsetzte, wären ihrerseits reflexiv[135]. Im Rechtssystem fällt man nicht die faktische Entscheidung selbst, sondern entscheidet nur über Entscheidungsprämissen. Kurz, unsere These ist: *Ein responsives Recht realisiert seine eigene reflexive Orientierung, indem es strukturelle Voraussetzungen für Reflexionsprozesse in anderen Sozialsystemen schafft.* Genau dies trifft das Verständnis einer integrativen Funktion des "responsive law".

## 4. Eine neue rechtliche Selbstbeschränkung: Entwicklungschancen im modernen Privatrecht

Sicherlich, das "Ergebnis" unserer bisherigen Überlegungen ist rein hypothetischer Natur. Daß responsives Recht sich in Richtung auf materiale und auf reflexive Rationalität hinentwickelt, und daß deren Potential sich unter Bedingungen einer post-modernen Gesellschaft nur unterschiedlich realisiert, dies sind Hypothesen, die wir durch selektive Ausbeutung dreier Gedankenkonstrukte aus unterschiedlichen Theoriekontexten gewonnen haben. Die Kombination "interner" rechtlicher Variablen (Theorie des "responsive law") mit "externen" gesellschaftlichen Variablen (Theorien der postmodernen Gesellschaft aus dem Kontext des Neo-funktionalismus bzw. der Kritischen Theorie) ergab ein Modell gesellschaftlich-rechtlicher Entwicklung, in dessen Rahmen die These aufgestellt wurde, daß "responsives" Recht nur dann auf die Herausforderungen einer post-modernen Gesellschaft angemessen antworten kann, wenn es sich in Richtung auf "reflexive" Rationalität entwickelt. Kein Zweifel, daß solch eine Hypothese, die auf spekulativen (und deshalb auf frag-würdigen) theoretischen Annahmen aufbaut, harte empirische Tests durchlaufen müßte, ehe sie irgendeine "Validität" beanspruchen könnte. Dennoch mag es sinnvoll sein, die Realitätsentsprechungen unseres Modells wenigstens zu illustrieren, dadurch, daß wir auf solche rechtliche Tendenzen hinweisen, in denen man "reflexive" Entwicklungschancen vermuten kann. Freilich muß

man bei einem solchen Vorgehen von vornherein einräumen, daß die "fallacy of misplaced concreteness" eigentlich unvermeidlich ist.

Vertragsrecht soll wieder als Beispiel dienen. Die bekannte Entwicklung der "socialization of contract" durch öffentliche Inhaltskontrollen drückt offensichtlich den Trend zu einer materialen Rationalität aus[136]. Gesetzliche Definitionen von Minimalbedingungen und richterliche Inhaltskontrollen sind die wesentlichen Kennzeichen einer Materialisierung des Vertragsrechts. Die Kapazitätsgrenzen dieser Vorgehensweise, in der staatliches Recht wieder die politische Verantwortung für unzählige Teilsystemordnungen übernehmen muß, werden heute sichtbar[137]. Ein Ausweg liegt nun möglicherweise in abstrakteren, indirekteren Steuerungstechniken des Rechts. Im kollektiven Arbeitsrecht sind solche abstrakteren Kontrolltechniken entwickelt worden, in denen sich "reflexive" Chancen ausmachen lassen. Ohne den Schwerpunkt auf konkrete Inhaltskontrollen zu setzen, tendiert dort das staatliche Recht dahin, kollektive Vereinbarungen nun dadurch indirekt zu steuern, daß auf die innere Organisation der Tarifverbände Einfluß nimmt, ihre rechtliche Anerkennung von bestimmten Strukturvoraussetzungen abhängig macht, Verfahrensnormen für das Verhandlungssystem und kampfweise Auseinandersetyung entwirft, die Kompetenzen der kollektiven Akteure ausweitet oder einschränkt. Dadurch wird nicht nur versucht, über eine Balancierung ihrer Verhandlungsmacht indirekt die Qualität der Verhandlungsergebnisse zu steuern. Wichtiger als die Neuarrangements von sozialer Macht mit Mitteln des Rechts sind Rechtsstrategien, die darauf abzielen, die "öffentliche Verantwortung" der konfligierenden Parteien im industriellen Konflikt zu erhöhen. Nicht gemeint sind hier die eher naiv anmutenden Versuche, durch explizite Normierung einer "Gemeinwohlklausel" den Tarifpartnern oder dem Tarifsystem als solchem eine öffentliche Verantwortlichkeit auferlegen zu wollen. Die Entwicklung von "Reflexionsstrukturen" im industriellen Konfliktsystem, von Strukturen also, in denen die gesellschaftlichen Folgen in Handlungskalkül aufgenommen werden, ist eher von einer institutionellen Beeinflussung von Organisationsgröße und inneren Organisationsstrukturen zu erwarten. Vergleichende empirische Studien stützen bis zu einem gewissen Grad die Vermutung, daß die gesellschaftsweiten Effekte innerhalb der Entscheidungsprozesse des Tarifsystems in einem größeren Ausmaß "reflektiert" zu werden pflegen, wenn das kollektive Arbeitsrecht ein zentralisiertes "Industriegewerkschaftssystem" im Gegensatz zu einem dezentralisierten "shop steward"-System mit berufsmäßig orientierten Gewerkschaften systematisch favorisiert[138].

Sicherlich kann das kollektive Arbeitsrecht kein Muster für eine universelle Anwendung im Vertragsrecht abgeben. Entsprechende Bemühungen auf anderen Gebieten, ein auf Gegenmacht basierendes Verhandlungssystem zu installieren, haben sich als nur bedingt entwicklungsfähig

erwiesen, so besonders auf dem Gebiet des Verbraucherrechts[139]. Jedoch mögen funktionale Äquivalente darin gefunden werden, daß mit staatlich-rechtlicher Hilfe "künstlich" autonome öffentliche Institutionen geschaffen werden, die Verbraucherinformationen liefern oder die Verbraucherinteressen für ihre wirtschaftlich-politische Repräsentation wirksam organisieren (vgl. als erste Ansätze in dieser Richtung z.B. Stiftung Warentest, Verbraucherzentralen, öffentliche Verbrauchervertretungen)[140]. Auch hier bestünde die Rolle des staatlichen Rechts nicht in materialer Regulierung von Marktprozessen, sondern in prozeduraler und organisationaler Vorstrukturierung von "autonomen" sozialen Prozessen. Über die Anordnung von Organisationsnormen zwingt das Recht hochspezialisierte, einseitig ausgerichtete Institutionen, widersprüchliche Anforderungen ihrer gesellschaftlichen Umwelten in das eigene Entscheidungskalkül aufzunehmen.

Die Konsequenzen einer solchen verbraucherrechtlichen Perspektive, die die Schwächen eines Rechtlichen Interventionismus überwinden könnte, werden von *Joerges* folgendermaßen gekennzeichent: "Das Recht hätte dann nicht autoritativ darüber zu befinden, was das Verbraucherinteresse 'ist'; es könnte sich damit begnügen, Kompetenzen für die Artikulation von Verbraucherinteressen zu bestimmen und deren Wahrnehmung zu sichern. Der Justiz fiele dann nicht die Aufgabe zu, eigene Zweckprogramme zu entwickeln, oder Zielkonflikte zwischen konkurrierenden Politiken zu entscheiden; sie könnte sich damit begnügen, Koordinationsprozesse zu gewährleisten und Einigungen zu erzwingen"[141].

Gerade das Verbraucherrecht, aber auch andere Bereiche des Vertragsrechts zeigen, daß die Technik der "externen Dezentralisierung", in der wir reflexive Entwicklungschancen vermutet haben, dann versagt, wenn soziale Macht- und Informationsasymmetrien die rechtliche Fernsteuerung von autonomen selbstregulatorischen Prozessen zur Farce machen. Kann das Recht dann in sich selbst "reflexive" Strukturen ausbilden, die kompensatorisch wirken? Eine etwas gewagte Vermutung geht dahin, daß solche "reflexive" Entwicklungschancen im Recht der Generalklauseln bestehen. Am Beispiel der Generalklausel von "Treu und Glauben" kann gezeigt werden, daß sie nicht notwendig als ein Instrument der "Verstaatlichung" des Vertragsrechts, also des richterlichen Interventionismus im Sinne eines materialisierten Rechts, interpretiert werden muß wie es gemeinhin geschieht, sondern eher als ein Instrument der "Vergesellschaftung" des Vertrages[142]. Damit ist gemeint, daß die Generalklausel dazu benutzt wird, die Abhängigkeit vertraglicher Erwartungsstrukturen von vielfältigen nicht-konsensualen gesellschaftlichen Steuerungsmechanismen sichtbar zu machen und diese vertragsintern zu koordinieren. Widersprüchliche gesellschaftliche Anforderungen werden von verschiedenen Ebenen an die Vertragsbeziehung gestellt – von der Interaktionsebene der Vertragspartner, von der Institutionsebene von Markt und Organi-

sation, von der gesamtgesellschaftlichen Ebene des Zusammenspiels von Politik, Wirtschaft, Recht. Diese Anforderungen zu kompatibilisieren bezeichnet die Konkretisierungsaufgabe der Generalklausel. Der "reflexive" Charakter eines solchen Vorgehens zeigt sich in der Notwendigkeit der rechtsinternen Simulation von autonomen gesellschaftlichen Prozessen. Auf der Interaktionsebene heißt dies, daß eine Verweisung auf informelle Verhaltenserwartungen in Frage steht, daß aber bei richterlich diagnostizierten "Interaktionsversagen" objektive Vertragszwecke und Verhaltenspflichten von Rechts wegen definiert werden. Entsprechend geht es auf der Institutionsebene darum, daß bei "Marktversagen" die bloße rechtliche Übernahme von "Verkehrssitten" und "Handelsbräuchen" ersetzt wird durch die Formulierung eines richterlichen Marktverhaltensrechts, durch die autoritative Formulierung von Aufklärungspflichten und professionellen Standards. Auf gesamtgesellschaftlicher Ebene verweist die Generalklausel auf die in Rechtsnormen verbindlich definierte "Politik des Gesetzes". Bei einem "Versagen" der politischen Mechanismen aber bleibt es der richterlichen Entscheidung überlassen, wie im Einzelfall die "public policy" zu formulieren ist. Die Logik dieses Vorgehens besteht darin, daß ein Defizit gesellschaftlicher Selbststeuerungsmechanismen in der Realität unterstellt wird und daß die Generalklausel die Anweisung enthält, solche *sozialen Selbststeuerungsprozesse rechtssystemintern zu simulieren.* Daß eine solche Simulation ihrerseits Rationalitäts- und Legitimätsdefizite aufweist, ja daß sie zum bloßen Moralappell an die billig und gerecht Denkenden verkommen kann, liegt auf der Hand.

Eine offene Frage ist es, ob man die Qualität solcher rechtsinternen Außenweltmodelle in kognitiver und in prozeduraler Hinsicht verbessern kann. Ein "Entdeckungsverfahren Praxis", verstanden als ein Kommunikations- und Lernprozeß zwischen "beteiligten Verkehrskreisen", Verbandsorganisationen, staatlichen Aufsichtsämtern, Ministerialbürokratien, akademischen Rechts- und Sozialwissenschaften, mag in dieser Hinsicht eine sehr praktische Entdeckung sein[143].

Gesellschafts- und Unternehmensrecht mag als ein weiteres Beispiel dienen. Auch hier finden sich Tendenzen zu einer Materialisierung des vormals neutralen Formenrechts. Doch auch hier scheinen staatliche Regulierung und richterliche Kontrolle des "corporate behavior" die Grenzen ihrer Kontrollkapazität zu erreichen, und das nicht nur wegen vordergründiger politischer Trendwenden, sondern aus strukturellen Gründen[144]. Unser Ansatz würde nun nicht zu einer Verstärkung staatlicher Machtressourcen führen, sondern eher zu einer Umstellung auf abstraktere Kontrollmechanismen: Rechtsnormen zu entwerfen, die systematisch "Reflexionsstrukturen" innerhalb des wirtschaftlichen Systems stärken. "Unternehmensverfassung" und "neokorporatistische Verhandlungssysteme" sind die einschlägigen Stichworte. Die offene Frage ist, ob sich solche Kontrollmechanismen dahin dirigieren lassen, daß sie als "corporate con-

science" effektiv fungieren, d.h. das Wirschaftssystem bzw. die Unternehmung dazu zwingen, gesellschaftliche Außenkonflikte zu "internalisieren". Und zwar in der Weise, daß interne Entscheidungsprozesse zugleich auch nicht-wirtschaftliche Interessen der Arbeitnehmer, der Verbraucher, der allgemeinen Öffentlichkeit berücksichtigen. Ist es völlig ausgeschlossen, daß ökonomische Zielstrukturen innerhalb der Unternehmung, die in den letzten Jahrzehnten schon erhebliche Wandlungen von Gewinnorientierung zu einer Wachstumsorientierung durchgemacht haben, sich erneut wandeln, so daß sie Probleme der ökologischen Balance in das Zielbündel effektiv mit aufnehmen[145]? Und muß dies wirklich der Punkt sein, "where the law ends", um mit *Christopher Stone* zu sprechen[146], sondern ist dies nicht gerade der Punkt, an dem Recht eine institutionelle Chance erhält: die "reflexive" Kontrolle wirtschaftlichen Verhaltens zu veranlassen, die externe soziale "troubles" in orgnisationsinterne "issue" für die Mikropolitik der Unternehmung zu transformieren geeignet ist?

Um es in diesem Zusammenhang zu wiederholen, die "Demokratisierung" sozialer Institutionen wandelt hier ihre traditionelle Bedeutung[147]. Weder Machtabbau noch Steigerung individueller Partizipation im emphatischen Sinne der "partizipatorischen Demokratie" sind die vorrangigen Ziele, sondern die überlegte Planung von internen Organisationsstrukturen, die die Institutionen – Unternehmen, öffentliche Verbände, Gewerkschaften, Massenmedien, Erziehungseinrichtungen – sensibel machen gegenüber den sozialen Effekten, die ihre Strategie zur Maximierung ihrer spezifischen Rationalität auslösen. Hauptfunktion solcher reflexiver Binnenstrukturen wäre es, interventionistische Staaatskontrollen durch effektive Binnenkontrollen zu ersetzen[148]. Strukturelle Bedingungen für ein "organizational conscience" zu schaffen, das die Balance zwischen Funktion und Leistung des Sozialsystems reflektiert, dies wäre in unserer Definition die integrative Rolle eines responsiven Rechts.

## 5. Einige Folgerungen für eine soziologische Jurisprudenz: Rechtliche Konstruktion sozialer Wirklichkeit

Diese Beispiele mögen die Bedeutung einer "reflexiven" Orientierung des Rechts illustriert haben. Schon aus ihnen wird ersichtlich, welche Relevanz einem ihrer Teilaspekte zukommt: der "kognitiven Kompetenz" des Rechts, die auch in der Theorie des "responsive law" einen zentralen Platz einnimmt[149]. Wenn die Annahme richtig ist, daß rechtliche Reflexion dadurch zu sozialer Integration beiträgt, daß sie zwischen Funktion und Leistung des Rechts vermittelt, dann sind die kognitiven Vorbedingungen solcher Selbstreflexion, besonders ihre "inneren Modelle der Wirklichkeit", in ihrer Bedeutung kaum zu überschätzen. Die Relevanz solcher rechtssystemeigener Wirklichkeitsmodelle – die sowohl empirische, prospektive als auch operative Elemente umfassen – wird in der Rechtstheo-

rie zunehmend anerkannt[150]. Nur als Beispiel mag *Dworkin* angeführt werden, der die Konstruktion "politischer Theorien" für notwendig hält, um in der Interaktion zwischen Normen und Prinzipien "hard cases" zu entscheiden[151]. Allgemeiner gesprochen lassen sich "Modelle" der sozialen Realität rekonstruieren, die Rechtsnormen, Gerichtsentscheidungen und dogmatischen Überlegungen regelmäßig zugrundeliegen, Modelle, die ihrerseits als kognitiver Hintergrund oder – um den hermeneutischen Ausdruck zu benutzen – als "Vorverständnis" dienen, um rechtliche Entscheidungen anzuleiten[152]. Modelle des Vertrags, der Verbandsorganisation, der Unternehmung, der Beziehung von Staat und Gesellschaft etwa gehören zu solchen spezifisch rechtlichen Wirklichkeitskonstruktionen, die ganz erheblich von unserem Alltagsverständnis, aber auch von wissenschaftlichen Theoriekonstruktionen abweichen. Solche Abweichungen lassen sich systematisch auf den jeweiligen sozialen Kontext zurückführen: Um soziale Konflikte unter normativen Gesichtspunkten rechtsförmig entscheiden zu können, muß sich das Rechtssystem im strengen Sinne des Wortes eine rechtseigene Wirklichkeit konstruieren[153]. In der Perspektive rechtlicher Konfliktlösung schafft sich das Recht buchstäblich seine eigene Welt, indem es hochselektive Modelle der Wirklichkeit abstrahiert und dabei viele seiner "relevanten" politischen, sozialen oder ökonomischen Elemente systematisch vernachlässigt.

Offensichtlich müssen solche Modellkonstruktionen mit der gesellschaftlich-rechtlichen Entwicklung zugleich variieren, und es stellt sich die Frage nach den Entsprechungen von Modelltypen und Typen rechtlicher Rationalität. Wenn schon Materialisierungsprozesse das Formalrecht gezwungen haben, seine Wirklichkeitssicht drastisch zu verändern, welche Art von Modellkonstruktion wird dann von reflexiver Rechtsorientierung gefordert?

Die soziologische Revolution im Recht (sociological jurisprudence, Freirechtsschule, Interessenjurisprudenz) sollte als das methodologische Gegenstück zu den Materialisierungsprozessen des Formalrechts gesehen werden[154]. Sie attackierte den Rechtsformalismus nicht nur wegen seiner "Begriffsjurisprudenz", wie es ihrem Selbstverständnis entsprach, sondern gerade wegen seiner rechtlichen Konstruktion der Wirklichkeit. Formale Rechtsrationalität hatte auf einer autonomen rechtswissenschaftlichen Konzeptualisierung der Welt bestanden; "ethische, politische oder volkswirtschaftliche Erwägungen" waren nach dem bekannten *Windscheid*-Zitat nicht Aufgabe des "Juristen als solchem"[155]. Demgegenüber führte die Rematerialisierung des Rechts zu einer Verwissenschaftlichung seiner Wirklichkeitsmodelle exakt in dem Sinne, wie "sociological jurisprudence" es forderte[156].

Ein solcher Typ soziologischer Jurisprudenz ist jedoch an materiale Rechtsrationalität auch in ihrer Krise gebunden, die wir im Kontext von Neo-funktionalismus und Kritischer Theorie analysiert haben. Wenn eine

solche methodologische Orientierung Ernst machen sollte, dann müßte sie in der Lage sein, hochkomplexe Wirklichkeitsmodelle zu entwickeln, die sozialwissenschaftliches Wissen in einem solchen Ausmaß voraussetzen und inkorporieren, daß das Recht die Verantwortung für umfassende Planungsprozesse übernehmen könnte. Juristische Analysen müßten dann in veritable "social policy"-Analysen überführt werden, die von einer adäquaten Situationsanalyse, über die Definition von Zielzuständen, die Auswahl von Normen als geeignete Mittel bis hin zu ihrer Implementation in der Sozialwirklichkeit führen müßten[157]. Offensichtlich würde die hierfür erforderliche Komplexität von Modellen, wenn dies Vorgehen Aussicht auf Erfolg haben sollte, die kognitive Kompetenz jedes heute existierenden Rechtssystems überfordern, selbst wenn es auf gründlichen interdisziplinären Analysen aufbaute[158].

Es scheint, als benötigte soziologische Jurisprudenz – parallel zu bestimmten Entwicklungen in der Entscheidungstheorie – ein Konzept der "bounded rationality", um arbeitsfähige Wirklichkeitsmodelle konstruieren zu können[159]. Wieder mag es die Rolle von rechtlichen Reflexionsprozesen sein, eine neue rechtliche Selbst-Beschränkung zu definieren, diesmal im Zusammenhang des juristischen Zugriffs auf Sozialwirklichkeit. Eine reflexive Rationalität des Rechts erforderte, daß das Rechtssystem sich als ein System in einer Umwelt begreift[160], um seine eigenen Kapazitätsgrenzen in der Regulierung anderer Sozialsysteme zu erkennen. Und dies erforderte ein Verhältnis von Rechtswissenschaft zu Sozialwissenschaft, das weder durch "Rezeption" noch durch "Abschottung" angemessen erfaßt wird, sondern nur als "Übersetzung" sozialwissenschaftlichen Wissens von einem Sozialkontext zum anderen nach bestimmten Übersetzungsregeln, d.h. nach bestimmten rechtseigenen Selektivitätskriterien[161]. Wenn unsere These richtig sein sollte, daß die integrative Aufgabe des Rechts darin besteht, Reflexionsprozesse in anderen Sub-Systemen systematisch zu fördern, dann ergeben sich hieraus bestimmte Beschränkungen für das erforderliche sozialwissenschaftliche Instrumentarium und für die rechtlichen Modellkonstruktionen. Beides unterschiede sich deutlich von einem umfassenden Planungsrecht materialer Rationalität. *Reflexives Recht müßte vorrangig soziales Wissen über selbstregulatorische Prozesse in unterschiedlichen Sozialkontexten verarbeiten – und neu entwickeln.* Umfassende Modelle einer rechtlichen "social policy" würden dann abgelöst von Modellen, die rechtlich-soziologische Analysen mit Interaktionsprozessen zur Lösung von Sozialproblemen kombinierten.

## Résumé/Summary

Droit réflexif: Les efforts les plus complets visant à développer une nouvelle approche de l'évolution du droit se trouvent dans les œuvres de

*Nonet* et *Selznick* aux états Unis et dans les travaux de *Habermas* et *Luhmann* en Allemagne. Ces théories neo-évolutionnistes identifient certaines phases de développment du droit, illustrent la progression d'une phase à l'autre et fournissent des futurs explicatifs des mécanismes de transition ainsi observés. Alors que toutes ces théories traitent d'un problème commun la crise de la rationalité formelle du droit, elles révèlent des différences sensibles lorsqu'elles abordent la question de l'émergence d'une rationalité post-moderne du droit. Dans cet article, l'auteur tente d'approcher ce phénomène en décomposant les modèles neo-évolutionnistes et en les restructurant d'une manière différente. C'est dans un modèle intégrant de la co-variation socio-légale que l'auteur identifie un type émergent de structures légales, nommé "droit réflexif". La conception est alors developpée par la mise en évidence des futurs expliquant comment la dynamique interne de l'évolution legale (*Nonet* et *Selznick*) est appelée à se développer dans le cadre des transformations sociales contemporaines (*Habermas* et *Luhmann*). La thèse centrale est qu'une orientation réflexive du droit constitue une nouvelle autorestriction du droit. Plutôt que d'assumer une responsibilité totale pour les effets des procès sociaux, le droit reflexif se restreint à l'etablissement, l'äménagement et la formulation nouvelle des mécanismes democratiques auto-regulateurs dans les subsystèmes sociaux.

*Reflexive law:* The most comprehensive efforts to develop a new evolutionary approach to law are found in the work of *Nonet* and *Selznick* in the United States, and *Habermas* and *Luhmann* in Germany. These neo-evolutionary accounts seek to identify different types of law, show the progression from one type to the other, and explain the processes of transition. While they all are concerned with a common problem – the crisis of formal rationality of law – they differ drastically in their account of an emerging post-modern rationality of law. The author tries to approach this phenomenon by decomposing the diverse neo-evolutionary models and restructuring them in a different way. In a comprehensive model of socio-legal co-variation, the author identifies an emerging kind of legal structure which he calls "reflexive law".

The development of this concept arises from the effort to explain how the internal dynamics of legal development described by *Nonet* and *Selznick* is likely to play itself out in the environment of societal transformatioon which form the basis of the theories developed by *Habermas* and *Luhmann*.

The central thesis is: A reflexive orientation of law leads to a new legal self-restraint. Instead of taking a regulatory responsibility for the outcome of social processes, reflexive law restricts itself to the installation, correction and re-definition of democratic self-regulatory mechanisms.

## Anmerkungen

[1] Vergleiche die Diskussion neuerer evolutionistischer Ansätze in der amerikanischen Rechtstheorie bei Lawrence Friedman, *The Legal System: A Social Science Perspective* (New York: Russell Sage, 1975), Kap. X, S. 283ff. (dt. *Das Rechtssystem im Blickfeld der Sozialwissenschaften*, Berlin: Duncker, 1981). Neuere deutsche Beiträge zur Evolution des Rechts: Klaus Eder, "Zur Rationalisierungsproblematik des modernen Rechts", *Soziale Welt* 29 (1978), S. 247ff.; Ernst Tugendhat, *Begründungsstrukturen im modernen Recht*, ARSP-Beiheft N.F. Nr. 14 (Wiesbaden: Steiner 1980); Andreas Zielcke, "Zur Rationalität des modernen Rechts", *Rechtstheorie* 11 (1980), S. 85ff.; Helmut Willke, *Zur Steuerungsfunktion des Staates in hochkomplexen Gesellschaften* (Köln, Typoscript 1981), besonders Kap. II.

[2] Philippe Nonet und Philip Selznick, *Law and Society in Transition: Toward Responsive Law* (New York: Harper, 1978).

[3] Niklas Luhmann, *Rechtssoziologie* 1 (Reinbek: Rowohlt, 1972) besonders Kap. III, S. 132ff.; "Evolution des Rechts", *Rechtstheorie* 1 (1970), S. 3ff.; "Subjektive Rechte: Zum Umbau des Rechtsbewußtseins für die moderne Gesellschaft" in *Gesellschaftsstruktur und Semantik* 2 (Frankfurt: Suhrkamp, 981), S. 45ff.

[4] Jürgen Habermas, *Zur Rekonstruktion des Historischen Materialismus* (Frankfurt: Suhrkamp, 1976); *Theorie des kommunikativen Handelns* 2 (Frankfurt: Suhrkamp, 1981), S. 257ff.

[5] Konventionell ist eine Moral, die durch Tradition gerechtfertigt ist, post-konventionell bezieht sich auf eine Rechtfertigung durch die Interessen aller Beteiligten, Jürgen Habermas, "Einleitung: Historischer Materialismus und die Entwicklung normativer Strukturen" und "Überlegungen zum evolutionären Stellenwert des modernen Rechts", in *Zur Rekonstruktion des Historischen Materialismus*, S. 9ff.; 260ff.

[6] Max Weber, *Rechtssoziologie*, (Neuwied: 2. Aufl. 1967), § 3, § 5, § 7, § 8. Amerikanische rechtssoziologische Interpretationen bei Lawrence Friedman, "On Legalistic Reasoning: A Footnote to Weber", *Wisconsin Law Review* 1966, S. 148ff.; David M. Trubek, "Max Weber on Law and the Rise of Capitalism", *Wisconsin Law Review* 1972, S. 720ff.

[7] Weber, aaO., S. 125.

[8] AaO., S. 331.

[9] AaO., S. 332ff.

[10] AaO., S. 125.

[11] AaO., S. 346.

[12] Besonders deutlich bei Eder, "Zur Rationalisierungsproblematik des modernen Rechts". Zu Materialisierungstendenzen im amerikanischen Recht vgl. Roberto M. Unger, *Law in Modern Society* (New York: Free Press, 1976), S. 192ff.; David M. Trubek, "Toward a Social Theory of Law: An Essay on the Study of Law and Development", *Yale Law Journal* 83 (1972), S. 1ff.; Gerald Turkel, "Rational Law and Boundary Maintenance", *Law and Society Review* 15 (1980-81), S. 41ff.; zur deutschen Entwicklung vgl. Franz Wieacker, *Privatrechtsgeschichte der Neuzeit* (Göttingen: Vandenhoeck & Ruprecht, 1967), S. 514ff. Neuere eingehende Analysen der Materialisierungsprozesse bei Rudolf Wiethölter, "Entwicklung des Rechtsbegriffs" in Gessner/Winter (ed.), *Rechtsformen der Verflechtung von Staat und Wirtschaft* (Opladen: Westdeutscher Verlag, 1982); Hans-Dieter Assmann, Gert Brüggemeier, Dieter Hart, Christian Joerges, *Wirtschaftsrecht als Kritik des Privatrechts* (Königstein: Athenäum, 1980).

[13] *Law and Society in Transition*, S. 78ff., 95ff.

[14] Luhmann, *Rechtssoziologie* 2, S. 325ff.; *Rechtssystem und Rechtsdogmatik* (Stuttgart: Kohlhammer, 1974), besonders S. 49ff.

[15] Habermas, *Legitimationsprobleme im Spätkapitalismus* (Frankfurt: Suhrkamp, 1973), bes. S. 120ff.

[16] Zur Fruchtbarkeit eines solchen Vorgehens vgl. Philip Selznick, "Jurisprudence and Social Policy: Aspirations and Perspectives", *California Law Review* 68 (1980) S. 209 mit Hinweis auf Maurice R. Cohen's "Principle of Polarity".

[17] *Law and Society in Transition*, S. 8ff.; Philipp Selznick, *Law, Society and Industrial Justice* (New York: Russell Sage, 1969), Teil I: "Legal and Social Theory", S. 3ff.; "Law. The Sociology of Law" in *International Encyclopedia of the Social Sciences* 9 (1968), S. 50ff.

[18] Vgl. die Kritik an verschiedenen rechtssoziologischen Ansätzen durch Luhmann, *Rechtssoziologie* 1, S. 1ff. und Helmut Schelsky, *Die Soziologen und das Recht* (Opladen: Westdeutscher Verlag, 1980), S. 77f.

[19] Eingehend zu dieser Kritik Luhmann, aaO.

[20] *Law and Society in Transition*, S. 20.

[21] S. 51f.

[22] Gustav Radbruch, *Vorschule der Rechtsphilosophie (Göttingen, 1965)*; vgl. auch Karl Engisch, *Einführung in das juristische Denken (Stuttgart: Kohlhammer, 7. Aufl. 1977), Kap. VI.*

[23] Vgl. die klassische Interpretation der Generalklauseln durch Franz Neumann, "Der Funktionswandel des Gesetzes im Recht der bürgerlichen Gesellschaft" (1937) in *Demokratischer und autoritärer Staat* (Frankfurt: Europäische Verlagsanstalt, 1967), S. 7ff.

[24] Max Weber, *Rechtssoziologie*, S. 332ff.

[25] *Law and Society in Transition*, S. 78ff., 95ff., 104ff.

[26] Habermas, *Zur Rekonstruktion des Historischen Materialismus*, S. 144ff.; *Theorie des kommunikativen Handelns* 2, S. 251ff., 257ff.

[27] Luhmann, "Evolution des Rechts".

[28] Der Begriff der Selbstreferenz wird sowohl in der Biologie als auch in den Sozialwissenschaften benutzt, um ein System zu identifizieren, das die Elemente, aus denen es besteht, produziert und reproduziert. Zu einer soziologischen Verwendung vgl. Niklas Luhmann, "Selbstreferenz und Teleologie in gesellschaftstheoretischer Perspektive", in *Gesellschaftsstruktur und Semantik* 2, S. 9ff.; in bezug auf das Rechtssystem "Subjektive Rechte", S. 100ff.; in bezug auf das politische System *Politische Theorie im Wohlfahrtsstaat* (München: Olzog, 1981), S. 33ff. Zu ihrer Interdependenz vg. "Machtkreislauf und Recht in Demokratien", *Zeitschrift für Rechtssoziologie* 2 (1981), S. 158ff.

[29] Gegenüber soziologischen Reduktionen betont gerade Selznick dieses Element, vgl. Selznick, "Law. The Sociology of Law", S. 50f.

[30] Vgl. zu dieser These Gunther Teubner, "Generalklauseln als sozio-normative Modelle", in Klaus Lüderssen u.a., *Generalklauseln als Gegenstand der Sozialwissenschaften* (Baden-Baden: Nomos, 1978), S. 13ff.; überarbeitete Fassung in Herbert Stachowiak, *Werte, Bedürfnisse und sozialer Wandel* (München: Fink, 1982), S. 87ff.

[31] *Law and Society in Transition* S. 78ff., 84, 87ff., 93.

[32] S. 84ff., 96, 106.

[33] S. 104ff., 108, 110ff.

[34] S. 16.

[35] S. 111ff.

[36] S. 95ff.

[37] S. 84ff.

[38] Habermas, "Überlegungen zum evolutionären Stellenwert des modernen Rechts", S. 262ff.; *Theorie des kommunikativen Handelns* 1, S. 349ff.

[39] Dies ist der Kern von Max Webers Begriff von formaler Rationalität, vgl. Max Weber, *Rechtssoziologie*, S. 123ff.

[40] Duncan Kennedy, "Form and Substance in Private Law Adjudication", *Harvard Law Review* 89 (1976), S. 1685ff., 1691.

[41] Luhmann, "Subjektive Rechte", S. 74.

[42] Habermas, "Überlegungen zum evolutionären Stellenwert des modernen Rechts", S. 264f.

[43] Luhmann, "Subjektive Rechte", S. 76ff.; vgl. Habermas, aaO., S. 262; entsprechende Analysen für die amerikanische Situation bei Lawrence Friedman, *A History of American Law* (New York: Simon and Schuster, 1973), S. 14ff.; Trubek, "Max Weber on Law and the Rise of Capitalism".

[44] Systematische Analysen dieses Zusammenhanges bei Assmann, Brüggemeier, Hart, Joerges, *Wirtschaftsrecht als Kritik des Privatrechts*, S. 32ff., 71ff., 187ff., 249ff.; Eder, "Zur Rationalisierungsproblematik des modernen Rechts"; Rüdiger Voigt (ed.), *Verrechtlichung* (Königstein: Athenäum 1980); Wolf von Heydebrand, "Der Einfluß des Staats-Wirtschafts-Komplexes auf die justiziellen Entscheidungsformen", in: Gessner/Winter, *Rechtsnormen der Verflechtung*.

[45] Als eindringliche Analyse, besonders in bezug auf Konsequenzialismus im Recht, vgl. Luhmann, *Rechtssystem und Rechtsdogmatik*, S. 31ff.; Thomas Wälde, *Juristische Folgenorientierung* (Königstein: Athenäum 1979).

[46] Manfred Rehbinder, "Status – Kontrakt – Rolle. Wandlungen der Rechtsstruktur auf dem Wege zur offenen Gesellschaft", in *Festschrift für Ernst E. Hirsch* (Berlin: Duncker, 1967), S. 141ff.

[47] Am Beispiel des Mitbestimmungsurteils des BVerfG und der Aussperrungsurteile des BAG, Wiethölter, "Entwicklung des Rechtsbegriffs", passim. In allgemeiner rechtssoziologischer Perspektive vgl. Willke, *Zur Steuerungsfunktion des Staates*, Kap. II. Zu einem neuen Prozeduralismus im amerikanischen Verfassungsrecht vgl. John Hart Ely, *Democracy and Distrust. A Theory of Judicial Review* (Cambridge: Harvard University Press, 1980).

[48] Eike Schmidt, "Von der Privat- zur Sozialautonomie" in *Juristenzeitung* 35 (1980), S. 153ff., 155f.

[49] Christian Joerges, *Verbraucherschutz als Rechtsproblem* (Heidelberg: Recht und Wirtschaft, 1981), S. 111ff.

[50] Luhmann, "Reflexive Mechanismen"; "Selbstthematisierung des Gesellschaftssystems" in *Soziologische Aufklärung* 2, S. 92ff., 72ff.

[51] Dazu Franz Lehner, *Grenzen des Regierens* (Königstein: Athenäum 1979), S. 178 ff.; Jürgen Gotthold, "Privatisierung oder Entbürokratisierung kommunaler Sozialpolitik" in Rüdiger Voigt, *Entrechtlichung* (Königstein: Athenäum 1982).

[52] "Law. Sociology of Law", S. 50, 55, 57; *Law, Society and Industrial Justice*, S. 243.

[53] *Law and Society in Transition*, S. 15, 21, 23, 116.

[54] Emile Durkheim, *Über die Teilung der sozialen Arbeit* (Frankfurt: Suhrkamp, 1977), Kap. 2 und 3. Vgl. für eine moderne Reformulierung von Durkheims Theorie der Rechtsentwicklung Jonathan H. Turner, "A Cybernetic Model of Legal Development", *Western Sociological Review* 5 (1974), S. 3ff.

[55] Vgl. als neuere Analyse der Rationalitätstypen in Max Webers Werk, Stephan Kalberg, "Weber's Types of Rationality: Cornerstones for the Analysis of Rationalization Processes in History", *American Journal of Sociology* 85 (1980), S. 1145ff.; grundlegend Wolfgang Schluchter, *Rationalismus der Weltbeherrschung* (Frankfurt: Suhrkamp 1980); Habermas, *Theorie des kommunikativen Handelns* 1, S. 225ff.

[56] Vgl. Habermas' "Sozialwissenschaftliches Konzept einer Krise" in *Legitimationsprobleme im Spätkapitalismus*, S. 9ff.

[57] *Law and Society in Transition*, S. 24, 33f., 115ff.

[58] Als amerikanisches Beispiel für diese Interpretation Unger, *Law in Modern Society*, S. 192ff.

[59] So die Interpretation von Habermas, "Legitimationsprobleme im modernen Staat", in *Zur Rekonstruktion des Historischen Materialismus*, S. 271ff.

[60] Habermas, "Geschichte und Evolution", S. 226ff.

[61] Habermas, "Moralentwicklung und Ich-Identität", in *Zur Rekonstruktion des Historischen Materialismus*, S. 63ff.

[62] Habermas, "Einleitung: Historischer Materialismus und die Entwicklung normativer Strukturen", S. 26ff.; *Theorie des kommunikativen Handelns* 2, S. 229ff.

[63] "Überlegungen zum evolutionären Stellenwert des modernen Rechts", S. 266.

[64] Habermas, *Zur Rekonstruktion des Historischen Materialismus*, S. 92ff., 129ff., 200ff., 260ff.; *Theorie der Gesellschaft oder Sozialtechnologie*, S. 270ff.

[65] Habermas, "Einleitung: Historischer Materialismus und die Entwicklung normativer Strukturen", S. 12.

[66] Habermas, "Zur Rekonstruktion des historischen Materialismus", S. 177ff.; "Geschichte und Evolution", S. 242ff.; vgl. auch *Theorie des kommunikativen Handelns* 2, S. 229ff.

[67] Habermas, *Theorie des kommunikativen Handelns* 2, S. 261. Vgl. zum folgenden "Überlegungen zum evolutionären Stellenwert des modernen Rechts", S. 262ff.

[68] Zur Evolutionstheorie vgl. Luhmann, "Evolution und Geschichte" und "Systemtheorie, Evolutionstheorie und Kommunikationstheorie", in *Soziologische Aufklärung* 2 (Opladen: Westdeutscher Verlag, 1975), S. 150ff. und 193ff.; *Gesellschaftsstruktur und Semantik* 1 (Frankfurt: Suhrkamp, 1980); "Differentiation of Society", *Canadian Journal of Sociology* 2 (1977), S. 29ff. Speziell zur Rechtsevolution vgl. Luhmann, *Rechtssoziologie* 1, S. 132ff.; "Evolution des Rechts"; "Subjektive Rechte".

[69] *Rechtssoziologie* 1, S. 136.

[70] "Evolution und Geschichte", S. 150ff.

[71] *Rechtssoziologie* 1, S. 140.

[72] *Rechtssoziologie* 1, S. 132ff.; "Evolution des Rechts", S. 7ff.

[73] "Evolution des Rechts", S. 16ff.

[74] Vgl. zusätzlich Luhmann, "Gerechtigkeit in den Rechtssystemen der modernen Gesellschaft", *Rechtstheorie* 4, (1973), S. 130ff., 142ff.; *Rechtssystem und Rechtsdogmatik*, S. 49ff.

[75] "Evolution des Rechts", S. 19.

[76] *Law and Society in Transition*, S. 29ff.

[77] Vgl. Habermas, "Zur Rekonstruktion des Historischen Materialismus", S. 176ff.; Luhmann, *Rechtssoziologie* 1, S. 166ff.

[78] Luhmann, *Rechtssoziologie* 1, S. 145ff.

[79] Luhmann, aaO. Vgl. für eine ähnliche Begriffsbildung Habermas, "Können komplexe Gesellschaften eine vernünftige Identität ausbilden?" in *Zur Rekonstruktion des Historischen Materialismus*, S. 97ff.; *Theorie des kommunikativen Handelns* 2, S. 261ff.

[80] Luhmann, *Rechtssoziologie* 1, S. 154ff.

[81] Habermas, "Zur Rekonstruktion des Historischen Materialismus", S. 176ff.

[82] Luhmann, "Differentiation of Society", S. 33.

[83] Habermas, aaO. In *Theorie des kommunikativen Handelns* 2, S. 246ff. löst Habermas dieses Organisationsprinzip in zwei Stadien auf: "hierarchisierte Stammesgesellschaft" und "politisch stratifizierte Klassengesellschaft".

[84] *Law and Society in Transition*, S. 29ff.

[85] Luhmann, *Rechtssoziologie* 1, S. 166ff.

[86] AaO., S. 171ff.

[87] Habermas, "Zur Rekonstruktion des Historischen Materialismus", S. 177.

[88] *Law and Society in Transition*, S. 51f.

[89] Habermas, "Geschichte und Evolution", S. 242.

[90] Luhmann, *Rechtssoziologie* 1, S. 190ff.

[91] *Law and Society in Transition*, S. 53ff.

[92] "Zum evolutionären Stellenwert des modernen Rechts", S. 262ff.; *Theorie des kommunikativen Handelns* 1, S. 349ff. und 2, S. 264f.

[93] Luhmann, "Positivität des Rechts als Voraussetzung einer modernen Gesellschaft", *Jahrbuch für Rechtssoziologie und Rechtstheorie* 1 (1970), S. 176ff.; *Rechtssoziologie* 1, S. 190ff.

[94] Nonet/Selznick, aaO., S. 70ff.; Habermas, "Geschichte und Evolution". S. 242; Luhmann, *Rechtssoziologie* 1, S. 190ff.

[95] Vgl. ihre Debatte in Habermas/Luhmann, *Theorie der Gesellschaft oder Sozialtechnologie*.

[96] Explizit Habermas, *Theorie des kommunikativen Handelns* 1, S. 334ff.; vgl. besonders sein "deskriptives Modell des Spätkapitalismus", das als ein Modell von Inter-System-Konflikten interpretiert werden kann, in Habermas, *Legitimationsprobleme im Spätkapitalismus*, S. 50ff. Vgl. auf der anderen Seite Luhmann, "Die Weltgesellschaft", in *Soziologische Aufklärung* 2, S. 51ff. Ein wichtiger Unterschied dürfte allerdings darin bestehen, daß Habermas solche Inter-System-Konflikte zusätzlich mit dem Dual "System/Lebenswelt" überformt.

[97] Habermas benützt explizit den Begriffsrahmen der Systemtheorie, vgl. nur *Legitimationsprobleme im Spätkapitalismus*, S. 9ff. Gegenüber solchen Übersetzungen würde Habermas jedoch darauf insistieren, daß damit eine befriedigende Analyse der "Lebenswelt" nicht geleistet werden kann, *Theorie des kommunikativen Handelns* 2, S. 171ff.

[98] Habermas, "Können komplexe Gesellschaften eine vernünftige Identität ausbilden?", S. 113ff.

[99] Insbesondere in Habermas, *Legitimationsprobleme im Spätkapitalismus*.

[100] AaO., S. 66ff.

[101] AaO., S. 50ff.

[102] Habermas, *Theorie des kommunikativen Handelns* 2, S. 530ff. Zur Materialisierung vgl. außerdem Brüggemeier, "Probleme einer Theorie des Wirtschaftsrechts", in Assmann, Brüggemeier, Hart und Joerges, *Wirtschaftsrecht als Kritik des Privatrechts*, S. 32ff., 71ff.; Unger, *Law in Modern Society*, S. 192ff.; Eder, "Zur Rationalisierungsproblematik des modernen Rechts".

[103] Habermas, *Legitimationsprobleme im Spätkapitalismus*, S. 73ff.

[104] Für das Verständnis von Habermas' Theorie ist es wichtig, daß sowohl die Rationalitätskrise als auch die Legitimationskrise des politischen Systems nur mögliche Krisen darstellen, während der dritte Typ, die Motivationskrise, einen notwendigen Konflikt zwischen dem politischen und dem kulturellen System aufdeckt.

[105] AaO., S. 99ff.

[106] AaO., S. 106ff.

[107] AaO., S. 131ff. In Habermas' neuester Veröffentlichung ist dies weit vorsichtiger – fast defensiv – formuliert. Nicht mehr die diskursive Durchdringung der Gesamtgesellschaft, sondern die Verteidigung kommunikativer Strukturen der Lebenswelt vor dem Zugriff der System-Rationalität steht jetzt im Vordergrund, *Theorie des kommunikativen Handelns* 2, S. 445ff.

[108] Habermas, "Legitimationsprobleme im modernen Staat", S. 217.

[109] AaO., S. 279.

[110] Habermas, "Zum Begriff der politischen Beteiligung" (1958) in *Kultur und Kritik* (Frankfurt: Suhrkamp, 1973), S. 9ff., 55ff.; *Strukturwandel der Öffentlichkeit* (Neuwied: Luchterhand, 1962), S. 228f., 269ff.; "Für ein neues Konzept der Hochschulverfassung", in *Protestbewegung und Hochschulreform* (Frankfurt: Suhrkamp 1969), S. 202ff.; "Verwissenschaftlichte Politik und öffentliche Meinung", in *Technik und Wissenschaft als "Ideologie"* (Frankfurt: Suhrkamp, 1968), S. 120ff.

[111] Habermas, *Theorie des kommunikativen Handelns* 1, S. 364.

[112] AaO., 2, S. 535ff.

[113] Vgl. speziell für die Rolle der Wirtschaft in der modernen Gesellschaft Luhmann, "Wirtschaft als soziales System", in *Soziologische Aufklärung* 1, (Opladen: Westdeutscher Verlag, 1970), S. 232ff.; für den Aspekt der Inter-System-Konflikte, *Rechtssoziologie* 1, S. 190ff.

[114] Luhmann, "Systemtheoretische Argumentationen. Eine Entgegnung auf Jürgen Habermas" in *Theorie der Gesellschaft oder Sozialtechnologie*, S. 375f.

[115] *Rechtssoziologie* 1, S. 190ff.; "Evolution des Rechts", Passim.

[116] *Rechtssoziologie* 1, S. 17.

[117] *Rechtssoziologie* 2, S. 325ff.

[118] *Rechtssystem und Rechtsdogmatik*, S. 31ff.

[119] Vgl. zu dieser Frage die eingehenden Analysen von Helmut Willke, "Zum Problem der Integration komplexer Sozialsysteme", *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 30 (1978), S. 228ff.; *Zur Steuerungsfunktion des Staates in hochkomplexen Gesellschaften*, Kap. II.; Turner, "A Cybernetic Model of Legal Development", S. 379ff.

[120] Durkheim, *Über die Teilung der sozialen Arbeit*, Kap. II und III.

[121] *Funktion der Religion*, S. 243.

[122] Luhmann, *Rechtssystem und Rechtsdogmatik*, S. 88, Fn. 108.

[123] "Differentiation of Society", S. 36ff.

[124] *Funktion der Religion*, S. 245.

[125] AaO., S. 54ff.

[126] "Differentiation of Society", S. 36f.

[127] AaO., S. 38.

[128] *Funktion der Religion*, S. 245ff.

[129] "Selbstreflexion des Rechtssystems", *Rechtstheorie* 10(1979), S. 176.

[130] Zu einer Ausarbeitung dieses Ansatzes vgl. Heinz-Dieter Assmann, *Wirtschaftsrecht in der Mixed Economy. Auf der Suche nach einem Sozialmodell für das Wirtschaftsrecht* (Frankfurt: Athenäum, 1980), bes. S. 122ff.

[131] Luhmann, *Rechtssoziologie* 1, S. 94ff.

[132] Vgl. zahlreiche Beiträge in Voigt, *Verrechtlichung*, die für verschiedene Politikbereiche die ambivalenten Folgen der Verrechtlichung untersuchen. Die Kritik ist aufgenommen und theoretisch verarbeitet bei Habermas, *Theorie des kommunikativen Handelns* 2, S. 522ff. und Luhmann, *Politische Theorie im Wohlfahrtsstaat*, S. 94ff.

[133] Habermas, aaO., S. 535ff.

[134] In der Argumentationstheorie ist dies besonders von Toulmin herausgearbeitet worden. Er differenziert Argumentationsstrukturen nach sozialen Handlungsfeldern, Stephen Toulmin, Richard Rieke, Allan Janik, *An Introduction to Reasoning* (New York: McMillan 1979). Für den subsystemspezifischen Fall des Rechtssystems selbst vgl. Tendenzen in dieser Richtung bei Robert Alexy, *Theorie der juristischen Argumentation* (Frankfurt: Suhrkamp, 1978), S. 259ff. Dies trifft sich mit parallelen Überlegungen im Rahmen der Modelltheorie, die sich mit systemspezifischer Selektivität von "inneren Modellen" befassen, vg. Herbert Stachowiak, *Allgemeine Modelltheorie* (New York: Springer, 1973), S. 40ff.; für den Fall des Rechtssystems Teubner, "Generalklauseln als sozio-normative Modelle", S. 13ff.

[135] Eine intensivere Diskussion der Begriffe Reflexion und reflexiver Mechanismen bei Luhmann, "Reflexive Mechanismen" und "Selbstthematisierung des Gesellschaftssystems"; "Identitätsgebrauch in selbstsubstitutiven Ordnungen, besonders Gesellschaften", in *Soziologische Aufklärung* 3 (Opladen: 1981), S. 178ff.

[136] Z.B. Wolfgang Friedman, *Law in a Changing Society* (Berkeley: University of California Press, 1959), S. 90ff., 106ff.

[137] Dies wird gerade von Autoren sehr deutlich gesehen, die selbst emphatisch für eine Deformalisierung des modernen Rechts plädiert haben, vgl. Wiethölter, "Entwicklung des Rechtsbegriffs".

[138] Wolfgang Streeck, P. Seglow, P. Bayley, *Railway Unions in Britain and West Germany. Structural Sources of Crossnational Differences and Similarities* (Berlin: Wissenschaftszentrum, 1978); Streeck, "Gewerkschaftsorganisation und industrielle Beziehungen", in

Joachim Matthes, (ed.), *Sozialer Wandel in Westeuropa*, (Frankfurt: Campus, 1979), S. 206ff.; Manfred Groser, "Stabilität und Wandel im System der Arbeitsbeziehungen", *Politische Vierteljahresschrift* 20 (1979), S. 258ff.

[139] Dieter Hart und Christian Joerges, "Verbraucherrecht und Marktökonomik" in Assmann u.a., *Wirtschaftsrecht als Kritik des Privatrechts*, S. 83ff.

[140] Vgl. Joerges, *Verbraucherschutz als Rechtsproblem*, S. 133, mit Hinweis auf entsprechende amerikanische Programme.

[141] Christian Joerges, "*Der Schutz des Verbrauchers und die Einheit des Zivilrechts*" (Typoskript 1982), S. 22f.

[142] Eine detaillierte Ausarbeitung dieses Ansatzes bei Gunther Teubner, in *Alternativ-Kommentar BGB*, Bd. 2, (Neuwied: Luchterhand 1980), § 242, 18ff.

[143] Dazu Joerges, aaO. und in Wendung auf das Deliktsrecht Gert Brüggemeier, in *Alternativ-Kommentar BGB*, Bd. 3, 2. Auflage 1982, vor § 823 C.

[144] Renate Mayntz, "Regulative Politik in der Krise" in Joachim Matthes, *Sozialer Wandel in Westeuropa.*

[145] Luhmann, "Differentiation of Society", S. 39.

[146] Christopher Stone, *Where the Law Ends: The Social Control of Corporate Behavior* (New York: Harper, 1975).

[147] Dies bezieht sich auf neuere Diskussionen in der Demokratietheorie. Als zusammenfassende Darstellung der Entwicklungen zu normativen komplexen Demokratietheorien, die die Kontroverse zwischen elitistischer und partizipatorischer Theorie überwinden, vgl. Gunther Teubner, *Organisationsdemokratie und Verbandsverfassung* (Tübingen: Mohr, 1978), Kap. 3.

[148] Zu einer Detailanalyse im Falle der öffentlichen Verbände vgl. Teubner aaO., Kap. 5 und 6.

[149] Nonet/Selznick, *Law and Society in Transition*, S. 78ff., 104ff., 112.

[150] Zur Modelltheorie im allgemeinen vgl. Lyndhurst Colling (ed.), *The Use of Models in Social Sciences* (Boulder: Westview Press, 1976); Herbert Stachowiak, *Allgemeine Modelltheorie*, (New York: Springer, 1973). In der Rechtstheorie vgl. Miguel Reale, "Pour une theorie des modeles juridiques", Acts of the IV. International Congress for Philosophy, 1970, Vol. 5, S. 144ff.; Teubner, "Generalklauseln als sozio-normative Modelle", S. 13ff.; Assmann, *Wirtschaftsrecht in der Mixed Economy*, S. 130ff.; für die Gesetzgebungswissenschaft Werner Maihofer, "Gesetzgebungswissenschaft" in Günther Winkler u.a. (ed.), *Gesetzgebung* (Wien: Springer, 1981).

[151] Ronald Dworkin, "Hard Cases" in *Taking Rights Seriously* (Cambridge: Harvard University Press, 1977), S. 81ff., 90ff.

[152] Josef Esser, *Vorverständnis und Methodenwahl* (Frankfurt: Athenäum, 1970).

[153] Peter Berger und Thomas Luckmann, *The Social Construction of Reality* (Garden City: Doubleday, 1966).

[154] Etwa Roscoe Pound, "The Scope and Purpose of Sociological Jurisprudence", *Harvard Law Review* 24 (1910/11), S. 591.; 25 (1911/12), S. 140ff., 489 ff.; "A Survey of Social Interests", *Harvard Law Review* 57 (1943/44), S. 1ff.; Philipp Heck, *Das Problem der Rechtsgewinnung. Gesetzesauslegung und Interessenjurisprudenz* (Bad Homburg: Athenäum, 1968).

[155] Bernhard Windscheid, "Die Aufgaben der Rechtswissenschaft" in: *Gesammelte Reden und Abhandlungen* (Leipzig: Duncker und Humblot, 1904), S. 101.

[156] Zur Verwissenschaftlichung des Rechts repräsentativ Hubert Rottleuthner, *Rechtswissenschaft als Sozialwissenschaft* (Frankfurt: Fischer, 1973).

[157] Vgl. die Kritik am "Rationalmodell" der Sozialpolitik bei Charles E. Lindblom und David K. Cohen, *Usable Knowledge. Social Science and Social Problem Solving* (New Haven: Yale University Press, 1979), S. 30ff.

[158] Luhmann, *Rechtssystem und Rechtsdogmatik*, S. 31ff.

[159] Herbert Simon, *Administrative Behavior* (New York: Free Press, 1976); Jomes G. March und Herbert Simon, *Organization* (New York: Wiley, 1966).

[160] Luhmann, "Selbstreflexion des Rechtssystems", S. 161ff.; Assmann, *Wirtschaftsrecht in der Mixed Economy*, S. 333ff.

[161] Teubner, "Generalklauseln als sozio-normative Modelle", S. 17f.; zu einer detaillierten Anwendung der modelltheoretischen Vorstellungen anhand konkreter Rechtsprobleme vgl. Rainer Walz, *Steuergerechtigkeit und Rechtsanwendung* (Hamburg: Decker, 1980).

# Supranationalism Revisited – a Retrospective: The European Communities after 30 Years

JOSEPH H. H. WEILER
Ann Arbor/Florence

## I. The thirtieth anniversary - a time for reflection?

The thirtieth anniversary of the European Community[1] does not provide a mere formal occasion for engaging in a retrospective analysis of the evolution of the main framework[2] for European integration. The early 1980s have been a focal point for several developments which are likely to affect fundamentally the future shape, direction and mode of operation of the Community. Three such developments deserve special mentioning.

(a) The accession of Greece[3] has opened the second phase of enlargement. Once completed the number of Member States originally parties to the Treaties of Paris and Rome will have doubled. The quantitative increase is likely to put serious strains on the Community's decision-making apparatus as well as raising a host of technical problems; the social and economic character of the new members is likely to put no less a strain on the substantive policies of the Community and the precarious balance of Member State interests which they represent.[4]

(b) The impact of the European Parliament – directly elected in the closing months of the last decade – on the institutional balance is likely to increase. The newly legitimized chamber has already flexed its muscles indicating its growing awareness of the discrepancy between its self-perceived functions and its constitutional powers. Institutional conflict is likely to increase.[5]

(c) The third development is a combination of several factors the coincidence of which goes beyond the periodic crises to which the Community has become accustomed.[6] The common agricultural policy, one of the mainstay Community policies, the elaboration of which in the 1960s was not only a major addition to the Community jurisdiction but also a political indication of the communion of interests between France and Germany, seems to be on a crisis course.

Not only are its effects criticized from without but even by its own internal budgetary rules the Community will not be able to finance it from

1982 onwards. The process of reformulating this basic policy is to take place in an economic climate of stringency, unemployment, energy shortages and balance-of-payment deficits which are likely to sharpen national differences and test the Community cohesion to its limit. The difficulties are already apparent with an ominous resurgence of protectionism. The commitment of the British Labour Party to withdraw from the Common Market and the Debré-Foyer Bill in the French Assembly are indications of strong grass roots feelings.[7] The French outright defiance of a European Court decision – the first incident of its kind in 30 years[8] – poses a much more direct and immediate threat, since if non-compliance became widespread, one basis on which the day-to-day operation of all Community policies depends will have been destroyed.

My purpose is not to try to predict solutions for these dilemmas nor even to examine the substantive issues which they raise. Rather, taking the 1980s as a turning point in the evolution of the Community, I shall attempt to analyse afresh the framework within which these and other problems will have to be resolved. My direct concern, through a retrospective analysis of legal and structural political developments and a prospective assessment of the future challenges to these developments, will be to give a clear picture of the special character of the European Community – still defying the traditional definitions of an international organization – as it appears today.

## II. Supranationalism - retrospective

### A - Supranationalism - An amorphous concept

The term traditionally employed in attempts to conceptualize the Community's politico-legal character – even if currently out of vogue – has been that of supranationalism.[9]

To define and give meaning and content to this term is to capture the essence of the particular model of integration experienced in Europe. The literal meaning of "... over and above individual States"[10] gives too general and antiquated a notion of supranationalism. For we know that whatever the dreams of the past, the Community's present system of governance involves "bits and pieces of the national governments..."[11] with a crucial say in all aspects of Community activity. But even so, in the transfer of certain functions to Community organs and in relation to certain constitutional hierarchies there remains a measure of "aboveness" in the old sense. Further, like all "federal" models[12] the Community presents a tension between the whole and the parts, centrifugal and centripetal forces, central Community organs and Member States. In some ways the balancing of this tension in the Community system departs – perhaps even

radically – from most other "federal" models. In the term supranationalism one must thus seek to give expression to these special phenomena; to map the "bits and pieces", to define the aboveness and to explain the European balancing of the "federal" tension. Yet trying to do so despite the passage of 30 years is an uneasy task. For although supranationalism is a term well established in the political and legal lexicon its usage is not unambiguous or without difficulty. It is possible to identify four interconnected issues around which these difficulties and ambiguities revolve.

### (a) *The definitional problem*

The need to revert to a novel term in characterizing the Community was indication that existing terminology was not felt to give adequate expression to the new venture.[13] But since the new term, supranationalism, derives from, indeed is the explication of, the phenomenon it seeks to define a measure of circularity is inevitable. For one must employ the distinguishing features of the Community to give meaning to the term supranationalism which in turn is the concept used to express distinctions between the Community and other international organizations. Thus, to the extent that the European Community experience was and remains unique one cannot usefully speak of a strict definition but only of one or more hallmarks which either identify it with, or – more importantly – distinguish it from other forms of association of States and legal orders.[14] In the earlier analyses the concept was tackled by reference to the known juridical-political categories of public international law (which embodies the formal traditional modes of international relations) and municipal law (which embodies the variety of non-unitary State arrangements).

In a powerful study of synthesis, Hay[15] analyses the most important of these attempts.[16] Few writers adopted extreme positions of total assimilation.[17] More often the analyses, using an analogical method, characterized supranationalism as *veering towards* one or the other of the two known systems. Insistence on finding a positive definition could only resolve itself in characterizing the system as *sui generis*. But to so characterize it with no more is hardly helpful.[18] And to say more leads inevitably to the necessity of fixing the distinguishing marks. Thus although the current trend is to include the Community among international organizations its distinguishing supranational character – whatever this may be – is always emphasized.[19]

Even if one cannot, as a result, have a fully fledged definition, the exercise of determining the most relevant hallmarks as indicia for supranationalism remains of interest. First, these may serve as a significant comparative tool for evaluating the similarities and differences between on the one hand both older and novel associations of States and on the other

hand the European Community. Also, should a true pattern of resemblance between other such associations and international organizations and the Community begin to emerge, the hallmarks may become a proper definition.

Secondly, in relation to the European experience itself, the hallmarks may provide a tool which will enable us to tackle the problem deriving from the dynamic nature of the Community and supranationalism.

(b) *The dynamic nature of supranationalism – the Community experience as a process*

In attempting to fix certain distinguishing features there is a danger of failing to give expression to the dynamic nature of supranationalism. The Community experience has not been a static one, neither in the substantive activities pursued by it nor – more significantly for our discussion – in the internal principles of its operation be they in the institutional framework, the location and exercise of decision-making power, and, generally, the relationships between the Community as a whole and its constituent parts. The interplay of centripetal and centrifugal forces – manifesting itself in different equilibria at different times – is a constant feature of the Community as a "federal" creature. Indeed, as will be readily accepted, in many ways the Community can only be understood as a type of dialectical process of action and reaction among the various forces shaping it. This poses a double-edged problem: one may select hallmarks reflecting the evolution of the Community at a fixed point in time. But with the passage of time these may be overtaken by events and cease to give a true reflection of supranationalism. Thus, for instance, studies which concentrated – as the primary distinguishing factor – on the ability of Community organs to take decisions immediately binding on individuals within the national legal systems[20] would, today, in the light of subsequent developments in the Commmunity, no longer reflect the mature evolution of that factor. Alternatively, a completely fluid set of hallmarks constantly changing with events would become no more than a description thereby losing its value as a comparative tool and its potential as an embryonic definition the importance of which was underlined above.

A compromise could be to shape as a dynamic tool a set of criteria within which the hallmarks would feature. The criteria would be sufficiently wide so as to be applicable to different legal orders and international organizations and so maintain their utility in comparative analysis. The hallmarks related to these criteria but changing with the evolution of the Community would provide the elastic element enabling us to understand supranationalism as a processual rather than fixed relationship or structure. This feature is particularly important in a retrospective analysis.

The processual character highlights, however, two further difficulties: the cleavage between legal and political assessment and the diffuse nature of the term supranationalism.

### (c) *Evaluating the supranational process – the cleavage between political and legal assessment*

Acceptance of the processual character of supranationalism and its subject, European integration, entails the possibility of evaluating it in the sense of establishing, with whatever measure of precision, the progress or retrogression of supranationalism and integration. It has been common to divide the process of European integration into "phases" and "periods" characterized by different degrees and levels of supranationalization and integration.[21] So as to avoid confusion in illustrating the legal-political cleavage it is necessary to draw a distinction between European integration and supranationalism. "European integration" is a concept wider than supranationalism for whereas the focus of the latter is on the constitutional, institutional and decision-making process within the Community, the former incorporates these processes but includes also the substantive developments in the areas covered by Community activities and their social and economic impact on the system as a whole. In this sense it could be said that supranationalism is concerned with the "means" and European integration with the totality of social, political and economic results. The separation is, naturally, not total since a strengthening of the means available for substantive integration, e.g. establishing that Community measures in general override national measures, may, in itself, be regarded as a substantive achievement. Thus, a diagnosis of stagnation in European integration as a whole would seem necessarily to entail a stagnation in the process of supranationalization. Conversely substantial progress of supranationalism would indicate progress in an important facet of European integration generally.[22]

The cleavage between the legal and political evaluation of the progress of European integration and supranationalism is most apparent by reference to contrasting assessments of different "phases" in the process.

The chasm appears both in the choice of phases and, more strikingly, in the evaluation of progress. Thus in a useful study synthesizing three decades of political theories on European integration, the period of 1958-69 is signalled out as a "distinct phase".[23] Evaluating this period the learned writer comments that "Throughout [these] 11 years during which General de Gaulle [who was "allergic to anything supranational"] remained in power, no notable progress could be made in integration, either in the political domain, the *institutional domain*, the monetary domain or in the geographical extension of the Common Market".[24] Yet, from the

juridical point of view – as shall be analysed in detail below – it was precisely during this period that certain fundamental facets of supranationalism took crucial, even revolutionary, strides ahead establishing, for example, the doctrines of direct effect and supremacy of Community law.[25] This, of course, is not a unique example of the cleavage.[26] To be sure the departure point of each discipline is different. The political theories of European integration were to a large measure wedded to a certain notion about the outcome of the process and embodied to a larger or smaller extent a certain predictive element about continuous progress. In addition, political theory laid great emphasis on the social, political and economic substantive achievements and lesser emphasis on means and ways. The starting point was thus one of high expectation and failure to maintain the visible social political momentum led to a measure of disillusion in the reassessment of European integration.[27]

The juridical point of departure was different. The constituent instruments of the Communities were traditional multi-partite international treaties which, even if including certain novel institutional features, were, in line with precedent, expected to be interpreted in accordance with the normal canons of treaty interpretation one of which is, for example, a presumption against loss of sovereignty by States. The process of integration which in the legal sphere was accomplished by a "constitutionalization" of the Treaties[28] was set against limited initial expectations. This conditioned among lawyers a far more positive evaluation of the process.[29] Besides, legal preoccupation especially in Europe has been concerned traditionally with means and tools, the legal instruments, and somewhat less with their political-social impact. Important as the supranational developments were, they formed only part of the total picture of European integration.

This attitudinal explanation cannot obscure the fact that in dealing with the same subject-matter, such a cleavage has emerged. Thus even in reviewing the narrower instrumental notion of supranationalism one of our concerns must be to try and provide some mechanism which will bridge, at least partially, this cleavage.

Finally, the existence of the cleavage within the context of the processual character of supranationalism, points to the last difficulty deriving from the complex nature of the term.

(d) *The "diffuse" nature of supranationalism*

In discussing the definitional problem we noted the need to rely on a multiplicity of distinguishing attributes – hallmarks – to give meaning to the term. We also noted the evolutionary nature of supranationalism reflecting the developments in the Community.

But how are these two factors to be related? The difficulty may best be illustrated by reference to the relationship between sovereignty and supranationalism.

Hay correctly suggests that "with few exceptions,... the criteria for the loss of sovereignty coincide with those which much of the literature regards as the elements of supranationalism.

Thus, the concept of a transfer of sovereignty may be the legal-analytical counterpart of the political-descriptive notion of supranationalism".[30]

The criteria which Hay distilled from the literature as elements of supranationalism include one or more of the following:

(i) The "independence of the organization and of its institutions from the Member States..."

(ii) The "... ability of an organization to bind its Member States by a majority or weighted majority vote".

(iii) The "... direct binding effect of law emanating from the organization on natural and legal persons".

(iv) The attribution to the organization of certain powers, functions and jurisdictions which in terms of sheer quantity result in a qualitative difference from non-supranational organizations.

(v) The nature of supranational institutions – principally the Parliament and Court.[31]

It is not necessary at this stage to examine in depth these criteria and evaluate the extent, if any, to which they fulfil a distinguishing function and meet the other problems mentioned above.

In general, in trying to choose the most apt hallmark one could of course focus on a single criterion such as the creation of a parliament or court as representing the most critical distinguishing factor and evaluate the process by reference to this single factor. The danger there would naturally be one of missing the complexity of supranationalism and indeed of European integration. The political-legal cleavage described above can, in some cases, be explained by an emphasis on one or some of the criteria to the exclusion of others by political scientists and jurists respectively. By selecting a multiplicity of factors a different danger is created – that of treating them as an *integrated* complex progressing or retrogressing as one whole.[32] A cursory examination of the list of criteria cited above will reveal that it is quite possible that in respect of some there may be "a loss of sovereignty" whereas in relation to others much less or none at all. How then is one to evaluate the progress or retrogression of supranationalism? To recognize the complexity of supranationalism which calls for a multiplicity of factors but to fail to realize that the processual character may take a different evolutionary direction in relation to each of these factors is to deny the diffuse nature of the term and to deprive the dynamic analysis of one of its important elements.

## B - Normative and decisional supranationalism

In order to overcome some of the difficulties outlined above it is submitted that a distinction should be drawn – if only for use as an analysis tool – between two facets of supranationalism. For convenience, I shall call them normative and decisional.[33]

Normative supranationalism is concerned with the relationships and hierarchy which exist between Community policies and legal measures on the one hand and competing policies and legal measures of the Member States on the other. Decisional supranationalism relates to the institutional framework and decision-making processes by which Community policies and measures are, in the first place, initiated, debated and formulated, then promulgated and finally executed.[34] The make-up of both these terms may include factors which can be stated at sufficient a level of generalization so as to serve as criteria for evaluating the progress or retrogression of the two particular facets of supranationalism.

A high measure of normative supranationalism will denote, in general, a hierarchy in which Community measures will take effective precedence over national ones. The choice of criteria is relatively simple since we may use the traditional principles by which relationships between on the one hand national and/or State law and, on the other, federal, confederal and international law are expressed: the principle of self-execution (direct effect), the principle of supremacy and the principle of pre-emption. The hallmarks will be the specific manifestation in the Community of these principles.

A high measure of decisional supranationalism will denote a process in which measures will be adopted and policies formulated and promulgated by means departing from traditional diplomacy and intergovernmentalism. Specifically in the Community context this will be indicated in relation to decisions taken:

(a) by Community organs the composition and mode of operation of which are autonomous (*communautaire*) and not intergovernmental in the traditional sense;

or

(b) by Community organs the composition and political functions of which are intergovernmental but the process of decision-making – e.g. the voting procedures – is not strictly that of intergovernmental diplomacy,

or

(c) in pluri-institutional decision-making where the role of the autonomous organs may be said to be critical, and in which the execution of these measures will be

(d) either directly by, or under the supervisory authority and responsibility of the autonomous bodies.[35]

Although all four factors are expressed in relation to the institutions of the Community, by virtue of the composition of these institutions, they contain the involvement, both direct and indirect, of the Member States in the decision-making process.

The separate treatment of these two facets covers the essence, if not the detail, of both the "juridical" approach which focuses on formal relationships, demarcation of competences and resolution of conflicts, and the more "political" approach which is concerned with the actualities, influenced by legal and non-legal factors, of cooperation and coordination of the different elements in the association of States. Having a clearer view of developments in both spheres may be helpful in drawing composite conclusions.

A final factor – relating to the Community's implied and additional powers – falls uneasily between the two facets and will consequently be treated separately.

Having introduced these distinctions it is now possible to analyse and trace the dynamic nature of supranationalism as a key to understanding the evolution of the legal-political framework of the Community.

### C - The dynamics of normative and decision supranationalism: diverging trends – and a resulting balance

Examination of the European Community's evolution in the last three decades reveals the apparently paradoxical emergence of two conflicting trends. One may have expected in the process of integration a parallel evolution in the transfer of power from the periphery to the centre – an increase in central normative competence accompanied by a strengthening in the "centralized" decision-making process. In the Community, however, we can trace on the one hand a, more or less, continuous process of *approfondissement* of normative supranationalism whereby the relationship between the (legal) order of the Community and that of the Member States has come to resemble increasingly a fully fledged (USA type) federal system. On the other hand, and contemporaneously, we can detect a, more or less, continuous process of *diminution* of decisional supranationalism, stopping, in some respects, only short of traditional intergovernmentalism. The very existence of supranationalism – in both its forms – in itself distinguishes the Community from most international organizations. The divergence in the evolution of the two forms may be one of the special – perhaps even unique – features of the Community as a process of integration and as a form of governance. The possible meaning to be given to these diverging Community trends will be discussed and assessed below. First, however, I will examine in greater detail the two facets of supranationalism with a view both to a clearer understanding of their meaning and so as to illustrate the respective evolutionary processes.

### 1. *Normative supranationalism: the process of approfondissement*

#### (a) SELF-EXECUTING MEASURES – THE DOCTRINE OF DIRECT EFFECT

The first distinguishing feature, or hallmark, of supranationalism in its early ECSC days was the power vested in the Community's main autonomous institution, the High Authority, to adopt self-executing measures which were directly binding on individuals – mainly undertakings in the coal and steel sectors.[36] Once the Treaty of Paris was ratified by the Member States this power could be executed *regardless of the monist or dualist character of the municipal legal order of the Member States*. Hitherto, traditional international organizations had the powers "... to negotiate agreements *ad referendum*; ... to take decisions which were binding on the members[37] but which would depend on national governments for their implementation... [and] to take decisions which the organization itself could implement".[38] The power of the Community's High Authority directly to bind individuals, subjects of national law, was, thus, a major innovation introduced by the Treaty of Paris which was acknowledged in most analyses of that period as the central characteristic of the Community. Despite the novelty of this feature, at least in modern times,[39] the international law character of the Treaty of Paris remained largely unaffected, since this manifestation of self-executing rule-making power was explicitly agree upon by the Member States which were signatories to the Treaty.

Later, however, this first characteristic of normative supranationalism was judicially developed in relation to the Treaty of Rome. In a series of landmark decisions the European Court of Justice, throughout the 1960s and 1970s took this doctrine far further than the limited provisions in the Treaties.[40]

It first held that subject to certain conditions, provisions of the Treaty of Rome – a Treaty which on its face resembled other treaties establishing international organizations – could become self-executing (have direct effect) bestowing enforceable rights as between individuals and the Member States.[41] Important as this celebrated decision may be in relation to the substantive consequences which would follow in respect of all Treaty Articles which could be shown to satisfy the conditions for direct effect, the main interest lies in the fact that this was the first major step in the "constitutionalization" of the Treaty of Rome[42] – its transformation by adopting a "constitutional interpretation" method into a quasi-constitution of the supranational entity. Implicit in the decision of the Court was the notion that the Member States were bound in their internal legal orders by their international treaty obligations. The individual was put, in certain respects, on a par with the State, a feature which usually appears only in municipal law. In other words, breach of international obligations, at least

those which were self-executing and materially capable of bestowing rights on individuals, became a matter of internal law.

Thus, Member States, *vis-à-vis* individuals, could no longer break their international treaty obligations relying on the weakness of traditional public international law. A weakness based in part on the exclusion of the individual as a direct subject of rights and duties (and of individual standing to sue) and on the traditional tardiness of States in bringing international claims on behalf of individuals when their national interest is not involved. The Court's ruling had another dimension since it gave a new vigilant and efficient guardian to international obligations – the individual.

Since that 1963 decision the doctrine of direct effect has been extended, deepened and elaborated. Important steps in its evolution have been its extension to create directly enforceable *Treaty* rights between individuals *inter se*[43] and its application, step by step, even to types of Community secondary legislation (e.g., directives) which are addressed to Member States and which on their face would not suggest the possibility of bestowing rights and duties on individuals.[44] The process of refinement continues to date.

### (b) THE DOCTRINE OF SUPREMACY

During the same period the European Court evolved its second crucial doctrine, the doctrine of supremacy, which, again, encapsulates a major aspect, in this sphere, of fully fledged federal legal systems. In another landmark case, *Costa* v *ENEL*,[45] the Court established a clear hierarchy of norms. In its view, which, according to the Treaty of Rome, is the authoritative view regarding the interpretation of that Treaty, Community law within the sphere of competence of the Community, be it primary or secondary, is superior to Member State law even if the latter is subsequently enacted and of a constitutional nature. As in the case of "direct effect" the derivation of supremacy from the Treaty depended on a "constitutional" rather than international law interpretation.[46] The Court's reasoning that supremacy was enshrined in the Treaty was contested by the Governments of Member States in this case and others. Acceptance of this view amounts in effect to a quiet revolution in the legal orders of the Member States. For, in respect of any matter coming within the competence of the Community,[47] the legal *Grundnorm* will have been effectively shifted, placing Community norms at the top of the legal pyramid.

It follows that the evolutionary nature of the doctrine of supremacy would – necessarily – be bi-dimensional. One dimension would be the elaboration of the parameters of the doctrine by the European Court. But

full reception thereof, the second dimension, would depend on its incorporation into the constitutional orders of the Member States and its affirmation by their supreme courts. It is relatively easy to trace the evolution of the Community dimension of the doctrine. In the *Costa* v *ENEL* decision, where it was launched, the Court was concerned with the paradigmatic conflict between substantive national and Community law. One may single out from the numerous cases in which it was affirmed the decisions in *Walt Wilhelm*[48] and *Simmenthal*[49] as illustrations of subsequent development. In the former, the Court accepted the possible legitimacy of having national competition policy operating besides the Community policy with each proceeding "... on the basis of the considerations peculiar to it".[50] Thus the issue was not about the possible co-existence of conflicting substantive law. But despite the legitimacy of having a parallel national competition policy which the Court did not dispute, the principle of supremacy required that a national court in proceedings before it in cases of national competition law must keep an open eye that its decisions even in their procedural, civil or penal aspects would not prejudice any concurrent Community – as yet incomplete – proceedings.[51] In the *Simmenthal* case the issue was not whether supremacy should exist but which court, in the national order, would decide this. The European Court, controversially,[52] but consistently with its earlier jurisprudence, insisted on the immediacy of supremacy so that even national procedural rules which did not deny the ultimate supremacy of Community law but designated internal procedures as to the court in which the review of the national legislation should take place, were prohibited.[53]

As regards the second dimension, the evolutionary character of the process is more complicated. It should be remembered that in respect of the original Member States there was no specific constitutional preparation for this European Court inspired development. The process of *approfondissement* may be thus seen in the gradual acceptance of the doctrine by the supreme courts of the Six. The pattern although uneven is clearly progressive. In some Member States the reception of the principle caused little problems,[54] in others, the Courts accepted the doctrine with reservations regarding the possible incompatibility of Community law with fundamental human rights enshrined in their constitutions.[55] Now that the European Court has indicated its willingness to review Community law itself in relation to a "higher law" of human rights based, in part, on the common constitutional traditions, these objections will have been somewhat quelled.[56] In others still, the judiciary split, with one branch accepting the doctrine and the other refusing it.[57] As regards the new Member States, especially those with a written constitution, the matter was simpler since at the time of accession supremacy was already an established principle and could be regulated formally in the process of accession.[58] The UK however presented a special problem since doubts remain as to the

very theoretical possibility of a shift in the *Grundnorm* of the type discussed above. The problem derives from the lack of written constitution and the conceptual difficulty of entrenching legislation – such as an Act giving prospective supremacy to Community law – so as to bind subsequent parliaments.[59] The matter is not fully resolved since no clear case involving UK legislation contradicting earlier Community law has come before the courts. In those cases in which the House of Lords occupied itself with Community law[60] it has, judiciously, avoided making a direct pronouncement on the subject. The Court of Appeal under the tutelage of Lord Denning has been less inhibited; its pronouncements – always *obiter* – see-sawed but have now settled on a halfway-house acceptance.[61]

So far we have treated the doctrine of direct effect and supremacy as distinct concepts, whereas analytically – linked by the Court's vision of the exigencies of a cohesive and integral legal order and its insistence on the principle of uniform interpretation and application of Community law – the two are tightly connected; in the sense supremacy is consequential of direct effect. Consideration of this connection will highlight another aspect of the evolutive nature of normative supranationalism. In *Van Gend en Loos* the Commission of the European Communities in its submissions stated that:

"... analysis of the legal structure of the Treaty and of the legal system which it establishes shows... that the Member States... [intended]... to establish a system of Community law and... that they did not wish to withdraw the application of this law from the ordinary jurisdiction of the national courts... [that Community law] must be effectively and uniformly applied throughout the whole of the Community. The result is... that the national courts are bound to apply directly the rules of Community law and finally that the national court is bound to ensure *that the rules of Community law prevail over conflicting national laws even if they are passed later.*"[62]

By contrast, the Advocate General in that case argued against the extension of direct effect to Treaty articles. He suggested that such extension – at a time in which the *principle of supremacy was not established* (and according to his exhaustive comparative analysis at least in some Member States Treaty law was decidedly not supreme over national laws) – would have the "... consequences of an uneven [non-uniform] development of the [substantive] law involved in the principle of direct application, consequences which do not accord with an essential aim of the Community".[63]

The Commission and Advocate General reached different conclusions in their submissions but were *ad idem* in seeing the inevitable linkage between supremacy and direct effect once the need for uniformity was established. It is submitted that the fact that the Court rejected this cue and preferred to introduce the two concepts into the Community legal order in two separate cases – even if, inevitably, using the very similar

"uniformity" argumentation – indicated a deliberate and politically wise attempt to phase out the progressive evolution of normative supranationalism so as to ensure as far as possible a smooth reception in the national legal and political orders. The strict connection between the two is evident also in *Simmenthal* where at times it is difficult to tell if the Court was applying the principle of supremacy or that of direct effect.

### (c) THE PRINCIPLES OF PRE-EMPTION[64]

It is here that one finds the third and final hallmark of normative supranationalism. In its purest and most extreme form pre-emption means that, in relation to fields in which the Community has policy-making competence, the Member States are not only precluded from enacting legislation contradictory to Community law (by virtue of the doctrine of supremacy) but they are pre-empted from taking any action at all. Initially, before the full ripening of the doctrine, the European Court achieved this objective by its earlier decisions which forbade the disguise of the Community nature of regulations.[65] This however was clearly insufficient. Subsequently, the principle came to its own even though it is still in an evolutionary stage. The Court of Justice is striving to attain an equilibrium between, on the one hand, the need to consolidate the *policy-making* capacity of the Community (which is at the essence of the pre-emption doctrine) and, on the other hand, the pragmatic necessity of regulation in fields in which the Community has competence but in which – for various reasons such as problems in its decision-making processes – it has not been able to evolve comprehensive Community policies. The Court has felt that in these situations the policy lacunae could be filled by Member State action implying a more flexible rendering of the pure pre-emption principle. In this the Court will be following in the footsteps of all federal systems none of which apply pure pre-emption. This shift in the Court's formulation may be illustrated by a number of decisions in the field of external economic relations. One of the questions in the *ERTA Case*[46] was whether the competence to negotiate and conclude an international agreement in the transport field rested in the Community or the Member States. Since, in its Transport Chapter, the Treaty does not refer specifically to Community competence to engage in international agreements, it was argued that such matters were to be left within Member State powers. In a judgment the importance of which goes beyond the specific question before us, the Court laid down an emphatic absolute principle of pre-emption:

"Each time the Community, with a view to implementing a common policy envisaged by the Treaty, adopts provisions laying down common rules, whatever form these may take, the Member States no longer have the right, acting individually or even collectively, to undertake

obligations with third countries which affect those rules. As and when such rules come into being, the Community alone is in a position to assume and carry out contractual obligations towards third countries affecting the whole sphere of application of the Community legal system."[67]

In the *OECD Case*[68] this emphatic statement was even strengthened whereby the Court said that the occupation of a field may be determined by the external act rather than by an initial internal measure:

"A commercial policy is in fact made up by the combination and interaction of internal and external measures... Sometimes agreements are concluded in execution of a policy fixed in advance, sometimes that policy is defined by the agreements themselves."[69]

In the case of *Kramer*[70] there is a certain retreat from the earlier emphatic position in *ERTA*. The Court stated that despite the fact that the Community had adopted internal measures, these "... limit themselves to providing the Community institutions with the power to take measures similar to those which the Member States... did take... The Community not yet having fully exercised its functions in the matter... the Member States had the power to assume commitments... and... the right to ensure the application of those commitments within the area of their jurisdiction".[71]

On the wide *ERTA* formula one may have expected absolute pre-emption and yet the Court in Kramer, mindful of the practical difficulties which such an approach might have had, was more lenient. This leniency was not, at that stage of the evolution of the principle, unqualified. The Court added that this concurrent Member States competence was transitional. The Treaties[72] had stipulated a deadline for the Council to adopt a fully fledged common policy and the prospect of adoption could have been an alternative explanation to the Court's pragmatic approach.

In the *International Rubber Agreement Case*[73] the Court accepted concurrent jurisdiction of the Member States in a field "occupied" by the Community "if... the financing (of the agreement in question) is to be by the Member States". The Court acknowledged however that as, unlike *Kramer*"... no formal decision has been taken on the question [as to whether the Community of Member States should finance the agreement, and as] there is no certainty as regards the attitude of the various Member States... The exclusive competence of the Community could not be envisaged in such a case".[74] In other words, the pragmatic approach was accepted even without the definite future prospect of a common policy. A pragmatic approach implies inevitably more difficulties in the determination of parameters and in the application thereof. Only through successive court decisions will these be clarified.

On its face then it would seem that in relation to the principle of pre-emption there has been a retardation rather than a deepening of the

scope of the principle in the Community legal order. For, in the terminology aptly adopted by Professor Waelbroeck, there seems to have been a shift by the Court from a "conceptualist-federalist approach" which corresponds to pre-emption in its purest and most exclusive form to a "pragmatic approach" which leaves the Member States concurrent competence with the Community. It should first be pointed out that in its second approach the Court has displayed, characteristically, a measure of political acumen. To insist on pure pre-emption when the Community institutions are not yet ready for their task could be retrogressive for the general evolution of the Community. At the same time the Court has insisted that in certain cases the pragmatic approach is transitory in nature. With the full occupation of a field by a Community policy, national measures (even if not contradictory and thus not in violation of the supremacy principle) could become prohibited *per se*. In other cases concurrent jurisdiction may remain constant. The *approfondissement* of pre-emption may thereby be explained in two ways. First, by its maturing from a crude even dogmatic statement of pure principle to a relatively sophisticated doctrine sensitive to Community needs. Secondly, pre-emption is seen to be spreading from one substantive field of Community law to another. It now affects sectors such as fisheries, competition policy and agriculture.[75]

### (d) The evolution of normative supranationalism – An interim assessment

The evolution of normative supranationalism has been both determined and rapid and, until recently, largely consistent. The moving force behind it, as we have seen, has been the European Court of Justice, the self perception of its role in the process of integration being the key to understanding the development. As a supreme adjudicator in a non-unitary system in which inherrent tensions exist between central institutions and the constituent member states, the Court of Justice had the choice between two different visions of its role. Article 164 EEC which charges the Court with ensuring "... that in the interpretation and application of [the] Treaty the law is observed" may indicate a vision in which the Court would be cast as an aloof and remote arbiter decidedly detached from the national-Community conflicts related to supranationalism and European integration. According to this vision the role of the Court could even be to prevent the normative evolution unless specifically agreed upon by the Member States. The entire jurisprudence of the Court of Justice represents a rejection of this approach, and, not surprisingly, it has come under acute criticism for this. Thus, in a critical even if sympathetic article, Hamson,[76] implicitly adopting the restrictive view based on Article 164, charged the Court in relation to the *Van Gend en Loos* case and its progeny of severing

"... the legal world – the world in which it operates – from the world of what we called real or actual events", of "short-circuit[ing] the scheme elaborated in the Treaty" and warned of the danger of the Court trespassing"... outside its province and [attempting] to establish by its own fiat what the Treaty directs to be established by a very different process".[77]

Although it is clear that the Court has not remained aloof in interpreting the Treaty and has taken a decidedly integrationalist approach, even Hamson has been careful not to charge the Court of actually overstepping the limits of the powers conferred upon it by the Treaty. The first question that has to be answered nevertheless is whether the integrationalist approach adopted by the Court is juridically legitimate within the scheme of the Treaty.

The key to answering this question and to understanding the Court's own vision of its role – which contrasts sharply with the narrow interpretation of Article 164 – must be found in the Preamble and Part One of the Treaty dealing with the principles upon which the Community is founded. The Preamble and Articles 1-3 declare the main objectives the attainment of which guided the founding fathers and the specific tasks and means to be pursued and employed in attaining these objectives. Thus we find, as the first provision in the Preamble, the objective of laying down foundations for an ever closer union among the peoples of Europe, in Article 2 the task, *inter alia*, of establishing a Common Market and in Article 3 the elimination of custom duties. For our purposes Article 4 is of importance. It provides that the tasks entrusted to the Community (namely the provisions indicated generally in the Preamble and outlined more specifically in Articles 2 and 3) shall be carried out by the Assembly (Parliament), the Council, the Commission and the Court, each institution acting within the limits of the powers conferred upon it by the Treaty. It is clear that the Court, within its powers and while faithful not only to the Treaty but also to the general fundamental guarantees of due process, is charged alongside the other institutions with forwarding the tasks entrusted to the Community. In *Van Gend en Loos*, the Court can be seen as taking its philosophy from Article 4 to which Article 164 is subordinate. *Van Gend en Loos* is a specific case which links the elements of uniting the peoples of Europe (Preambule), establishing a common market (Article 2) and eliminating custom duties (Article 3) in one maverick doctrine which simultaneously achieved all three effects. The entire pattern of decisions creating the normative framework represents a *grand design* to achieve the same effect.

In the light of Article 4 the Court has not stepped outside the province entrusted to it. It has, in a committed manner, fulfilled its task as perceived by the Treaty. This, however, does not preclude us from pointing out the implications of this perception according to which the Court "regards itself as the trustee of the hopes and aspirations, the

purposes and the objectives of the founders of the Community...".[78]

The Court was able – at least for a time – to maintain the momentum of normative evolution because of the traditional insulation which separates courts from *direct* pressures from the other political actors. At the same time the very vision adopted by the Court – a vision which, in the wide sense, gives it a measure of partizanship in the Community – Member State tensions – has the possible effect of eroding the traditional insulation and diminishing the authority with which the Court may speak. This in turn may lead to a curtailment of the Court's power to preserve and even continue the process of normative evolution. This inter-relationship between the judicial process and the political process will be discussed in the light of the analysis of decisional supranationalism.

### (e) Limits to jurisdiction?

Direct effect, supremacy and pre-emption are the core attributes not only of supranationalism, but can be found also in fully fledged federal systems. Whence then, the acknowledged paramountcy of the Member States in the Community system?

First, it should be recorded that even today the Community remains, functionally, a *multi-sectoral condominium*.[79] The Community legal order, with all its above attributes, extends only to those fields for which the Community has competence. But this alone would not be sufficient explanation. Several developed federal systems are based on a doctrine of enumerated powers which[80] however did not prove an obstacle to subsequent wholesale expansion. Indeed the evolution of European integration from the limited spheres of the Coal and Steel Community to an "Economic Community" indicated a great increase in the number and breadth of these sectoral condominiums. Further, the common market itself has seen a continuous process of extending the limits of Community competence often through the Community's own variant of judicially created doctrines of additional and implied powers.[81] The impact of Community law and the range of Community policies extend in some cases far beyond that which a literal reading of the Treaties might have suggested. One may perhaps talk here of "substantive *approfondissement*". Nevertheless, there are many fields, some even critical such as defence, fiscal and monetary policy and education, to mention just a few, which remain – for the time being – outside the Community sphere.

As regards the Community's expanding jurisdiction, although the very existence of Article 236 EEC, which specifies the method of Treaty amendment, indicates that theoretical limits do exist to this expansion, from the normative point of view the European Court of Justice – true to its integrationalist ethos – has consistently attempted to narrow those

limits interpreting widely and creatively existing provisions so as to allow jurisdictional expansion without recourse to complicated Treaty amendment.[82] In this, it has somewhat emulated the jurisprudence of another court – the US Supreme Court – in the latter's treatment of, say, the USA Commerce Clause.[83] However, the expansive approach adopted by the European Court of Justice is explicable, at least partially, by the different political structure of the Community, a structure which highlights in another way the centrality and paramountcy of the Member States. For in Europe, unlike the USA, national governments are responsible, by and large, for decision-making at the national level and at the Community level. When policies are adopted, even at the jurisdictional limits of the Treaty and beyond, the central decision-making organs of the Member States will have partaken in the process, thereby diffusing the debate on *ultra vires* competences.[84]

This then is another clue to Member State centrality. For the principles of normative supranationalism provide only a framework into which substantive rules and policies must be fitted. The Community system did not opt for a classical federal governmental structure characterized by a federal legislature directly elected "by the people", and a federal executive likewise selected, both bodies separate and independent of their would-be national counterparts. Instead the Treaty provides for a system in which the Member States governments play a key role in "filling in" the normative framework. However, the Fathers of the Treaty hoped that by creating a hybrid structure of decision-making which would differ substantially in its operation from traditional intergovernmental organizations, the Community interest would prevail despite the prominence of the Member States. Decision-making was also to be supranational. We must turn then to examine the evolution of decisional supranationalism.

## 2. *The diminution of decisional supranationalism*

Decisional supranationalism and its expression in the evolution of decision-making in the Community are, by comparison to normative supranationalism, less easy to trace and analyse. Several reasons account for this difficulty. Strangely (or, perhaps, wisely), the Treaties are rather cryptic in their institutional provisions. A literal reading of texts gives little indication as to the function of the institutions and only a formal indication as to their competences and powers. Inevitably, there is an enormous gap between these formal provisions and the actual *Realpolitik* manifestation of power in Community life. We noted, in discussing normative supranationalism, the preeminent role played by the European Court of Justice in widening and deepening the scope and meaning of normative supranationalism. Apart from all other features the judicial process is characterized by

a high measure of transparency which facilitates the task of the Community observer. It was possible to identify with relative ease and precision the evolving stages of normative supranationalism. By contrast the process of political decision-making and policy formulation is much more obscure[85] and its evolution is marked less by clear-cut landmarks – although some critical ones exist – and more by a subtle process of institutional interplay.[86] The tension between "whole and parts" is, naturally, a constant feature of this field as well. It manifests itself here in two, sometimes converging, axes: (1) Community versus Member States; and, within the Community, (2) non-intergovernmental versus more intergovernmental institutions. It would, naturally enough, be far too simplistic to suggest that decision-making may be explained by simple reference to these axes. The formulation of Community policies is a complicated and multi-phased process and the duality of axes manifests itself at almost each stage.[87] This will be illustrated below. At the same time, if a global view is adopted – one which would correspond to that which was adopted in relation to normative supranationalism – it is possible to detect a clear enough evolutionary line in decisional supranationalism – namely its decline. This decline is apparent in relation to all criteria which were used to characterize decisional supranationalism.

(i) The independence and the autonomous policy and decision-making role of the intergovernmental institutions is declining;

(ii) the weight of non-intergovernmental institutions in pluri-institutional decision-making processes is declining;

(iii) within quasi-intergovernmental institutions there is a decline of their unique supranational features;

(iv) in the execution of Community policies there has been a shift to Member States domination.

To understand this decline we must first discuss briefly the political institutions themselves and then turn to the decision-making process.

The main European Community institutions are sufficiently well known and do not need detailed description here. In the first three decades of Community life, apart from the Court, the institutions which dominated the scene were, clearly, the Commission (and High Authority) and the Council of Ministers. Also clear enough is the general role assigned to both organs. The Commission and its staff – although reflecting the national composition of the Community[88] – is undoubtedly the more *communautaire* and less intergovernmental of the two. The Commissioners are required by the Treaty "in the performance of [their] duties [neither to] seek nor take instructions from any government or from any other body... Each Member State undertakes to respect this principle and not to seek to influence the members of the Commission in the performance of their tasks".[89]

The Commission as a body is thus notionally autonomous from the

Member States and specifically free to pursue the Community interest. The functions of the Commission are varied. It has the exclusive role of initiating legislation (in the formal sense); according to this scheme it is also the central administrative organ of the Community;[90] it has a potentially important "diplomatic" role in acting as broker between the Member States; it acts as "federal agency" in those spheres where the Community has assured fully fledged federal powers (e.g. competition policy); it supervises the execution of the Treaties and Community law by Member States and acts as a "supranational public attorney general" in case of violation. According to early theory of the Community it was to be the technocratic "functional" core which would engineer and precipitate the famous spillover leading the Community towards political union.

The Council of Ministers by contrast is the main formal legislator – clearly indicating the centrality of the Member States. It would be wrong, *stricto sensu*, to characterize the Council as an intergovernmental institution. Its rules of voting and procedure, the role of the President and the formal reliance on Commission draft legislation distinguish it from classical intergovernmental organs and characterize it as supranational as well – albeit in a limited way.[91] Also, one important feature must be emphasized if the supranational character of the Council is to be fully appreciated. It is true that the Council can refuse to enact any legislation put before it by the Commission even if the policy is clearly in the interest of the Community. In rejecting such legislation, or amending it,[92] the Council can be – and often is – motivated by interests which are contrary to the "Community spirit".[93] The Member States are, however, obliged to act jointly, though not necessarily unanimously, within the framework of the Council of Ministers. This is one of the striking interactions of normative and decisional supranationalism. In any given field which is regulated by Community law and for which competence has been transferred to the Community organs, the individual Member States are precluded from taking unilateral action in implementing and/or changing the policy; they must act jointly as the Council of Ministers within the normal decision-making procedures of the Community.

Thus, in the field of agriculture the Community has evolved a common policy which determines *inter alia* annual price rises. The Council is free to reject any proposal from the Commission or from one of its members. Any member may veto any given proposal and thus block a decision. This is an indication of the low measure of decisional supranationalism in the Council. What is, critically, not permissible is for one Member State – or more – to "go it alone" in the face of Council deadlock. That is the expression of high measure of normative supranationalism. If decisions in the particular field are to be taken, they must be taken by the Ministers acting *qua* Council. At least to that limited extent the Council of Ministers remains a supranational institution. In theory, then, the tandem of Commission –

charged with policy initiative, a secondary legislative[94] function and with execution and supervision tasks – and Council – charged with policy decision-making and actual "primary" legislation thereby representing directly the interests of the Member States – was meant to achieve the balance in the decision-making process between Community and Member States. The real story has been very different, its main theme being the ever increasing strengthening in the weight of the Council and Member States and a corresponding decline of the Commission.[95] This process has manifested itself in several ways:

### (a) The decline of the Commission: the signs

In the early years of the Coal and Steel Community, the High Authority enjoyed a large measure of autonomy. Its responsibilities were fairly narrow – confined to these two sectors – and its functions resembled the current Commission functions in, say, the field of competition. Its main addressees were sectoral undertakings and the governmental departments directly concerned. This relative narrowness explains perhaps its measure of autonomy. The effect of High Authority activity on the Member States was *ex hypothesi* rather limited. Politically it did not emerge as a serious focal point of real power. Moreover, normative supranationalism was as yet fairly embryonic. The High Authority was king in its court.

With the conclusion of the Treaty of Rome the sphere of activities of the Communities received an enormous qualitative and quantitative boost. Despite the deliberate attempt to play down the supranational character of the Community, since 1958 Community decisions have had a much greater effect on national life: establishment of the customs union, the pursuit of the Communities' four "freedoms" (free movement of goods, labour, services and capital) and the creation of Community common policies inevitably and increasingly encroached on national policies, national legislative competences and national administrative freedom. One could expect an *a priori* greater interest and greater involvement of the Member States in the Community process. A further problem was bound to develop – the emergence of the Community "democracy deficit".[96] The democracy deficit has two tightly connected aspects. In the first place given the possible, often exaggerated, tensions between Community interests and national interests there was a fear that Community programmes would be developed which did not pay due respect to national interests. The European Assembly (Parliament) was denied by the Treaties any meaningful say in the evolution and supervision of policies and law. Its members were not directly elected to their posts but nominated by and from national parliaments. The Commission for its part lacked any direct popular legitimacy and the Council of Ministers – the main legislative body – represented the executives of the Member States. It was only to

be expected that uneasiness about the Community legislative process would develop. In the absence of an effective democratically legitimate *Community* check on the legislator, it was inevitable that there would be a shift to *national* scrutiny. Alas, national parliaments have created largely unimpressive mechanisms for control of Community actions.[97] Instead a powerful Council of Ministers which was, at least in theory, answerable to national parliamentary control could be seen as some answer to this aspect of the democratic deficit.[98] This solution is largely illusory. Much of the Council work is done by the civil servant Coreper members which are no more "legitimate" than the Commission and with the possible exception of Denmark, there is little direct control, except on major issues, on ministerial work within the Council. Indeed – and this is the other aspect of the deficit, which remains unanswered to date – the national ministers may use the legislative forum of the Council so as to pass legislation which may have been checked and even vetoed in the national parliaments. The other traditional check on the legislature was of course to be the Court of Justice. In the early 1960s, however, the Court saw as its main task the evolution and consolidation of European institutions and policies. Thus, for example, when a Community measure allegedly violating fundamental rights enshrined in a constitution of a Member State was brought before it for judicial review, the Court adopted a narrow formalistic approach and declared in effect that the Treaties did not impose a duty to respect those rights.[99] The Court was concerned not to impede in any way the function of the Community institutions nor to threaten the primacy of Community law. Later the Court learnt that the process of integration would be enhanced rather than impeded by a bold policy of judicial review.[100] The Court's early attitude added thus to the democracy deficit.

Despite these two factors, the enormous increase in the range of Community activities and competences, and the emergence of the democracy deficit – both of which would seem to suggest an inevitable increase in the importance of the Council of Ministers – the eclipse of the Commission was not immediately apparent. Two inter-related factors contributed to the maintenance of Commission power. First, the proximity in time to the conclusion of the Treaty of Rome bestowed legitimacy on the activities of the EEC in its first years. The Treaty and its provisions were debated in all six national parliaments and were approved. Secondly, so long as the Community was seen to be confined to an implementation of the explicit operative parts of the Treaty, the democratic deficit did not come to the fore. Given that the Commission's main task was the execution of these explicit policies – principally the establishment of the customs union – it did not come into major policy conflicts with the Council. (There were in fact certain personality clashes, in itself indication of the political self-perception of the Commissioners and their President in the early years).[101]

The Commission was thus able to gain immense prestige by the rapid and professional manner in which it coordinated the process of implementation and by its skill in fulfilling its "broker" role in securing the agreement of, and settling the disputes among, the governments of the Member States. Once these first tasks were substantially achieved, however, the process of erosion in the Commission's position began to become more transparent.

The signs of the Commission decline are clear enough and discussed ably elsewhere.[102] Of the more important signs one should mention the rise to eminence of the Committee of Permanent Representatives (Coreper) as a powerful intermediary between the Commission and Council and the initial exclusion and subsequent toleration of the Commission in the new policy-shaping European body – the European Council. The European Council was set up *hors* the Treaties and became the institutionalized forum for the meeting of Heads of State or Government. Its function was not only to shape new "second generation" policies but also to serve as the main arena for the settlement of intergovernmental disputes regarding Community issues. The rise of Coreper meant that Commission initiatives were subject to intergovernmental influence at an extremely early stage in their formulation thereby detracting from the Commission role as representing the *Community* vision.[103] The emergence of the European Council contributed to the detraction from the Commission's role as a source of Community fresh ideas and as the broker between the Member States.[104] The Commission was hemmed in by two new bodies "usurping" both its technical and political-diplomatic functions.

### (b) THE COUNCIL OF MINISTERS – EROSION OF SUPRANATIONAL FEATURES: THE SIGNS

The erosion of decisional supranationalism has not only been apparent in the decline of the Commission *vis-à-vis* the Council (in both its derivatives: European Council and Council of Ministers). Within the Council of Ministers itself there has been a decline in its supranational characteristics. We have alreay noted the emergence of the European Council. This, it is submitted, is an indication of the failure of the Council of Ministers' "First Eleven" – the Foreign Ministers – to assert themselves as an institutional body capable of giving direction to the Community and solving its problems. The need to resort to old style loosely structured summitry is a clear regression in the role of the Council of Ministers *qua* Community body.

The second, perhaps more, important landmark in the decline of decisional supranationalism was the retreat by the Council of Ministers from majority voting – which had been designed as the clearest manifes-

tation in the decision-making process of the precedence of the Community interest over the national interest – to consensus decision-making. This move, precipitated by France and at first grudgingly accepted by the other five in the legally dubious Accord of Luxembourg, was to become, with the accession of the three new Member States,[105] an accepted Community norm. Thus, one of the truly outstanding supranational features of the Council's procedure was reduced. Majority voting itself is not entirely exceptional in international organizations. It is the law-making power of the Council and the effect of that law on and in the national legal orders as expressed in the concept of normative supranationalism which made the prospect of majority voting so unique. The existing veto power which each Member State now holds does not necessarily paralyse the Council because, as we have seen, in areas controlled by the Community the Council *must* at the end of the day take a decision if entire policies are not to come to a halt. The power to veto does not give an individual Member State the power to impose its own desire as to the eventual outcome of the decision-making process. Rather the effect has been to force the nine partners into "package-deal decision-making" with compromises being sought not only as regards each policy but among various policies. This development was no doubt instrumental in the emergence of the European Council as a forum for this high powered political horse-trading although, as submitted above, the Council of Foreigh Ministers could have assumed this function.

The Luxembourg Accord and the power of veto did not completely destroy Community superiority in decision-making. First, there is the simple fact that, albeit by consent of the Council, many issues are still decided by majority voting.[106] Secondly, the veto power of the Member States in itself does not necessarily entail, as most commentators assume, the blocking of the Community supranational process. What is crucial is the legislative context in which the veto is exercised. When the Treaty provides explicitly for Council unanimity in adopting certain policies,[107] the veto power thereby entailed gives the individual Member State the ability directly to block any unacceptable measure. And to the extent that the Luxembourg Accord extends this power to measures in which the Treaty provides for majoritarian or qualified majoritarian voting, the same ability will naturally exist.

However, Article 149 EEC provides that when acting on a proposal from the Commission "unanimity shall be required for a [Council] act constituting an amendment to that proposal". In this case, then, the veto power available to the individual Member State gives it the power to prevent any tampering with a Commission proposal – i.e. a "supranational veto" – but not the ability to force an amendment. The only way open to a recalcitrant Member State whose proposed amendments are "supranationally blocked" by the veto of another State insisting on the Commission original proposal is to veto the entire measure which of course

is a much more serious matter and demands a higher threshold of national interest. Indeed Parliament made astute use of this principle in its latest budgetary wrangle with the Council. The ability of three Member States to block by a "qualified veto"[108] the changes which Council wished to introduce to the 1980 Parliamentary supplementary budget enabled that budget to be adopted by the President of Parliament.

### (c) Decline of decisional supranationalism – The execution of policies

The strict doctrine of separation of powers according to which the executive is concerned solely with implementation of policies adopted elsewhere was probably never tenable and the crucial policy-making role of executives is so apparent as to obviate any analysis. The definition of the executive function in the Community system is not easy. We have already noted that one traditional function of executives in contemporary Western democracies, that of initiating and submitting policy proposals to the legislative branch – a task initially associated with the Commission – has, except in the narrow technical sense, been taken over to a large extent, by the Council of Ministers and the European Council.

Has there been a similar decline in the post-legislative phase? The issues here are complex and our conclusions must be regarded as rather tentative and speculative. In some limited field, such as competition, the Commission acts like a federal agency with full executive powers although even here it has to rely for enforcement measures on the national systems. In most matters, however, the practical execution of Community policies and rules, be they in the field of agriculture, external imports and the like, is performed by the national administrations acting as "agents" for the Community in such things as collecting charges, issuing clearances and dispensing grants. To the extent that the "agency" is automatic, acting directly on Community measures and Commission instructions, this feature cannot be viewed as a weakness in the executive role of the Commission and as a sign of decline in decisional supranationalism.

However, the Communities as a political system have not escaped the world-wide trend of increased government by *administrative* legislation and action. Thus, whereas the legislature (the Council) enacts enabling measures or issues policy mandates, it is left to the executive to implement the policies by series of secondary legislative measures and administrative acts. Just as the Coreper was introduced by the 1965 Merger Treaty as an extra tier of Member State representation designed ostensibly to facilitate the technical preparation of Council meetings but in practice precipitating an erosion in the policy initiatory role of the Commission, it is possible to identify a similar trend in the establishment of Member State committees

in relation to the executive functions of the Commission. I shall focus, by way of illustration, on two aspects of this development: first, by way of micro-analysis on the role of one committee already provided for in the Treaty of Rome, and, secondly, by way of overwiew on the general proliferation of committees set up by Council secondary acts.

(i) The "Article 113 Committee"

Article 113 EEC dealing, *inter alia*, with conclusion of trade agreements by the Community bestows responsibility on the Commission for *negotiating* agreements although acting on a mandate given by the Council. The Commission must consult with "... a special committee appointed by the Council to assist the Commission in this task...". This "113 Committee" was duly set up,[109] composed of representatives of the Member States; its presidency held in rotation by the Member States holding the presidency of the Council. Its mode of operation is a useful illustration of the diminishing executive role of the Commission.[110] On the one hand a Council-Member State Committee would be useful especially if amendments to the mandate were needed and could be provided at the *locus* of negotiations. The committee would also ensure that agreements negotiated would receive the ultimate assent of the Council upon which the constitutional power to conclude the agreement is bestowed. At the same time it has been pointed out that the role assigned to the committee (and to national observers in other types of Community agreements) has meant that "... the Member States are able to oversee the Commission's behaviour in every negotiating session and insure that its application of the mandates to concrete issues comports with their wishes"[111] and that thus the "... picture of the Commission... is that of Community spokesman and agent or, more technically, plenipotentiary of the Council". The term "negotiate" has clearly not been interpreted [by virtue of the role assumed by the Member States and the Council] to accord the Commission an authoritative role in forming EEC negotiating policies or *directing the negotiations themselves*.[112] To the extent that the Commission has managed to re-build its role it has not relied on a reinterpretation of the Treaty provisions but on the sheer technical expertise which it can bring to the treaty-making process. But in this sense it is no more than an equivalent to a competent national civil service.

It is perhaps worth pointing out that use of a "democracy deficit" argument to justify these developments is hardly convincing for, except in the most formal sense, curtailment of Commission power has reduced the role of one set of (European) civil servants and elevated another (national) set. The continued reluctance of the Council to increase the role of the European Parliament in the process of treaty-making[113] testifies to the fact that national interests play a greater role than does concern for democracy.

(ii) The proliferation of committees

With the substantive expansion of Community activities, especially in the agricultural sector, a wide ranging network of committees has been set up, many of them to partake in the "legislative implementation" of Community policies. Some of these are advisory committees the consultation of which is obligatory but the opinion of which once consulted is not binding.[114] Others, the management and regulatory committees, have a more decisive role.

As regards the management committees, particularly prominent in the agricultural and fishery fields, the Commission must submit its draft measures to them. The committee may approve the measure – by a qualified majority – or fail to reach a decision (if no qualified majority is reached either way) whereupon the Commission is free to adopt the measure which will have full legislative force. If the committee manages in fact to "reject the measure" (qualified agreement against) the Commission may still adopt the measure but this measure will come into effect only after a certain period in which time the full Council may – by qualified majority – reject it. The regulatory committees procedure is slightly more restrictive on the Commission.[115]

It is difficult to assess this proliferation of committees in relation to the power of the Commission. Unlike, say, the "113 Committee", the Commission presides in all these management and regulatory committees. And since executive power is exercised here through legislative means the involvement of the Council would seem to be natural. Pragmatically, the involvement of representatives of the Member States may also contribute to smooth functioning of whatever policy is executed regardless of the implications to the institutional balance. Besides, in the decision-making process the qualified majority rule means that – as regards management committees – the vetoing power of a few States may actually assist the Commission in adopting the measure. All these elements would tend to point to the conclusion that the proliferation of committees serves the ends of open and efficient government.[116]

At the same time – taking a longer term view – one cannot avoid noting that these mechanisms indicate an unwillingness of the Council and Member States to entrust the execution of policy to the Commission with, say, safeguards of reporting and information. In this sense, then, the proliferation of committees may be regarded as one element in the decline of decisional supranationalism.[117]

(d) THE REASONS FOR DECLINE

It now remains to try to give some reasons for this erosion in the position of the Commission and the general decline of decisional supranationalism. Several such reasons may be given.

(i) We noted the early success of the Commission in implementing the explicit policies in the Treaties. The need to evolve a "second generation" of Community policies based on broad indications in the Treaties but not explicitly set out imposed a much more delicate and politically sensitive task on the Commission. The power of "initiative" now called for was less formal and technical, that is it no longer called for proposals which gave legislative form to explicit Treaty obligations but rather it called for wide, reflective and more "value" prone proposals. Whereas the first generation of policies were negative – in the sense that the Member States undertook to refrain from certain actions and the Commission was charged with implementing this negative regime – the new policies were to have an actual positive content on which agreement was more difficult. The ability of the Commission to propose new initiatives and get them accepted was thus considerably weakened. The importance of the Council of Ministers and European Council was strengthened. It was those bodies which could give the drive to new policies. To be sure, the Commission retained its position as a source of ideas and vision for developments on the Community level. The environmental policy, to give but one example, would not have emerged without an internal Commission initiative. But it was for the European Council to sanction the policy, to balance it and officially to launch it. The Commission was left somewhat in the background.

(ii) The need for "second generation" policies brought the democracy deficit to the fore. What is more, with the widening of Community activities, national parliaments felt threatened by a process which would wrest even more power from them. The Commission which had little formal democratic legitimacy became an easy target for attack.

(iii) The Commission itself, although growing in experience, put on much bureaucratic fat. This expressed itself not only in its numerical staff growth but also in the evolution of traditional bureaucratic ailments which significantly reduced its internal efficiency.[118] Lack of internal (lateral) coordination, distortions in the pattern of promotion, and personal and national rivalries all contributed to an internal drop in morale and an external drop in esteem. The materially privileged position of Commission employees (alongside all other Community employees) in the current harsh climate may also have contributed to this process.

(iv) The independence of the Commission and its judicial impartiality began breaking down in a bizarre dialectical process. The very importance of the Commission prompted the Member States to take not only – as was constitutionally permissible – an active interest in the appointment of the Commissioners themselves but also in the promotion of personnel within the Commission and in the allocation of portfolios among Commissioners. The Commission was thus seen to be losing its impartiality, a singular handicap when considering its intended mediatory role in Council disagreements. Internally, this governmental intervention contributed

both to a strain on the collegiate nature of the Commission and the authority of its President.[119]

(v) Finally, it may be that the process of *approfondissement* of normative supranationalism, as described above, had a negative effect on decisional supranationalism both in the Council-Commission relationship and within the Council itself. Normative supranationalism meant that the impact of Community policies and law was perceived as growing not only in scope – to cover more fields – but also in depth – so as to have a more immediate and binding legal effect from which the Member States could not escape. Thus, the politically delicate issue of supremacy was countered by an insistence of the Member States on their control of the making of this "supreme" law and their ability to block its making. This view of the role and power of national governments was emphasized strongly by "pro-marketeers" in the 1975 UK Referendum, using it (somewhat misleadingly) as a tool against charges of the "loss of sovereignty" which Community membership entailed. It is thus suggested that the correlation between the *approfondissement* of normative supranationalism and the diminution of decisional supranationalism is not accidental but at least partially causal. Thus as we noted, the European Court of Justice had not excluded the sensitive field of foreign economic relations from the effects of normative supranationalism – giving one of its decisive rulings – the *ERTA* decision – on the doctrine of pre-emption in relation to the treaty-making capacity of the Community. The Council of Ministers and the Member States, for their part, have ensured that in the execution of the Community's external policy the Commission is kept – by devices such as the 113 Committee – on a very tight rein and that the Member States, by a legal interpretation of the limits of Community competence, are active parties in many agreements.[120]

### 3. *The diverging trends – An assessment*

It is this last reason which gives the clue to one possible significance of the diverging trends. For, it is submitted, the outcome of the process represents a certain balance of action and reaction, whereby the permeation and expansion of Community influence – expansion in breadth expressed by the growing number of fields where Community impact is felt and expansion in depth as expressed by the *approfondissement* of normative supranationalism – is matched by an ever closer national control exercised in the decision-making processes. To the extent that the *approfondissement* and *diminution* are causally connected the relationship is surely two way : a cyclical interaction of the judicial-normative process with the political-decisional one. Here then is one dimension of the Community formula for attaining an equilibrium between whole and part, centripetal and centrifu-

gal, Community and Member States. It is an equilibrium which explains a seemingly irreconcilable equation : a large, surprisingly large, and effective measure of transnational integration coupled at the same time with the preservation of strong – unthreatened – national Member States. Schonfield, in his seminal political analysis of the Commmunity,[121] noted this duality juxtaposing on the one hand "... the extent to which the detailed operation of the Community powers is today jealously observed and controlled by the member governments"[122] with, on the other hand, two European Court cases[123] which demonstrated the high level of what I have called normative supranationalism. He created the "bag of sticky marbles" metaphor to express this duality and, correctly, sought the explanation in the *substantive* features which both draw and repel the members of the Community : A Community characterized on the one hand by a shared, historical, political and cultural background which, since repeatedly threatened by strife and conflict in the past and finding common economic and political exigencies in the present, was and is being pushed and drawn towards integration. But also a Community which is, on the other hand, equally characterized by a rich diversity which has evolved in a history of separate tribal, relegious, linguistic, economic and political development confinement of which into a fully fledged federal framework is neither feasible nor desirable.[124] The analysis of normative and decisional supranationalism complements the substantive explanation of the "bag of marbles" by giving it its *instrumental* expression. It also gives us a clue about the future of European integration. The divergence between normative and decisional supranationalism indicates that many of the tools for integration in the normative armoury of the Communities – even the more subtle tools, like the directive – are likely to remain underemployed because of decisional difficulties. In the preparation of new policies and programmes the Commission will have to shift emphasis. In present conditions it is no longer sufficient to introduce grand programmes for Community policies, based on a reasoned presentation of Community needs and potentials. In relation to each policy the Commission will also have to engage in a micro-analysis of the barriers to Council decision-making, which may vary from issue to issue, with a view to presenting programmes particularly sensitive to decisional obstacles as well as to normative ones. In this context the discussion of committees takes a different turn. For what could be seen as a reduction in the Commission's autonomous executive role could be usefully employed, by that very Commission, as a device to *reduce* the decisional barriers. The greater the involvement of the representatives of the Member States, the less reluctant may they be to new Commission proposals.

The Commission can, in certain fields, radically reassess its own role. Where the decisional barriers seem insurmountable it could abandon, selectively, its function as initiator of policies and engage in direct-volun-

tary-contact with the Member States to assist them to develop independent policies which may however serve Community ends.

Finally, relating the two facets of supranationalism – while recording that the distinction between the two was made so as to give us an additional tool of analysis – provides us with a different set of terms of reference, albeit a tentative set, with which to examine the jurisprudence of the Court of Justice and the type of integrationalist philosophy which, as we saw above, was at the basis of its judicial policy. The Court's case law and judicial policy may be subjected to four different orders of analysis.

The first – technical-legal order is concerned with ascertaining the precise meaning of parameters of the Court's judicial decisions. Thus, for instance, whereas the case law pertaining to supremacy seems consistent and clear there remains, as regards the doctrine of direct effect, especially as it applies to directives, a measure of obscurity with which the Court is currently grappling.

The second order of analysis is concerned with evaluating, in the strict legal sense, the correctness and legitimacy of the Court's case law. The challenge of the French Conseil d'Etat in the *Cohn-Bendit Case* as regards alleged constraints which Article 189 imposes on the European Court is a judicial reflection, even if feeble, of this order of analysis.

The third order of analysis is concerned with the evaluation of the Court's case law as regards its impact on European legal integration. It hardly needs repeating that without the bold steps taken by the Court, the most important of which were discussed briefly in our analysis of the evolution of normative supranationalism, the legal structure of the Community would manifest only a fraction of its present cohesion. So, unlike the second order of analysis the terms of reference here go well beyond a strict legal evaluation of the legitimacy of the Court's judicial policy. They are widened to incorporate the impact of these decisions on the general architecture of the legal order. But even here the analysis is still confined to an area which flows almost directly from the Court's jurisprudence. The evaluation, to use our terminology, remains at the level of normative supranationalism.

The fourth order of analysis then will try to widen the terms of reference even further. The questions which will be asked – and to which, at this stage of research, only extremely tentative answers may be given – concern the relationship between the Court's case law and its impact in the normative field on the one hand and the decisional facet of supranationalism on the other. This is, of course, a two-way relationship. I have already suggested that in evolving its doctrine of pre-emption the Court will have been cognizant of decisional difficulties in the Communities' policy and rule-making structure. To insist on pure pre-emption and expect it to work necessitates efficient central organs; the absence of these

in the Community gives one explanation to the pragmatic – less pure – approach adopted by the Court in this instance. It was on the basis of this analysis that I had argued that what appeared as a retrogressive judicial approach was in fact part of the *approfondissement* process. One could at this stage only speculate whether the Court would have taken a different approach as regards pre-emption had the decline in decisional supranationalism not occurred. Looking at the relationship the other way round one may speculate whether the ongoing process whereby the Court is, say, giving a higher normative value to directives (e.g. construing even non-directly effective directives as a basis for judicial review of Member State legislation) – a process motivated undoubtedly by the integrationalist approach of the Court – may be one of the reasons for the decisional difficulties in adopting certain directives. Even if one cannot expose a direct causal link, the mere possibility should certainly feature in the minds of the European judicial policy maker.

## D - Towards an "all or nothing effect"

There remains one crucial factor missing from the instrumental analysis. What are the ties that keep the framework together? By what means has it been possible to take normative supranationalism to the degree that it was taken despite the evident decline in decisional supranationalism, the outright hostility from certain national quarters[125] and the lack of independent federal enforcement mechanisms the evolution of which one would normally expect to accompany the development of normative supranationalism?

### 1. *Withdrawal or selective application*

The natural departure point for a reply would be an examination of the possibilities of Member State withdrawal from the Community. Here again we find if not a cleavage between legal and political analysis at least a sharp difference of emphasis. Juridically, in discussing withdrawal from the Community a distinction is drawn between the European Coal and Steel Community on the one hand and the European Economic Community and Euratom on the other. As regards the former, Article 97 ECSC provides that "This Treaty is concluded for a period of 50 years...". As regards the latter, Articles 240 EEC and 208 Euratom provide respectively that "This Treaty is concluded for an unlimited period". A recent study[126] based on general international institutional law,[127] cogently argues "... that no right of withdrawal can be implied in the case of treaties like ECSC...".[128] The same view is reached as regards the EEC and Euratom, namely that the aforenoted Articles "... exclude any implied possibility of

unilateral withdrawal" since "... no other meaning which could be given to these provisions would not make them redundant".[129]

This juridical argument is curtly dismissed in a political analysis of the Community system: Pryce argues that "The question of unilateral withdrawal is not dealt with in any of (the Treaties) – for the good reason that none of the partners was willing to tie its own hands with regard to this matter... The silence on this means quite clearly that each of the signatories maintains absolute authority to take such a decision at any point in the future. The fact that the Treaty of Paris was concluded for a 50-year period and the other two for an unlimited period is irrelevant in this context".[130]

This legally dubious statement has a strong measure of political truth behind it. Should a Member State be determined to withdraw, lack of legal consent from its partners will not be an obstacle in its way. The general Hobbesian maxim that covenants without swords are but words is accurately reflected by Pryce in his statement that in the Community "there is no army to convince a reluctant partner".[131] The real glue that binds the Community together is the bond of common vision and common interest in pursuing what has aptly been called "alliance politics".[132]

Does this mean that the supranational system is irrelevant to the discussion of withdrawal ? It is submitted that the question of unilateral withdrawal is the wrong one to ask. Unlike pre-World War II practice it is rare in the life of contemporary international organizations for States to withdraw their membership.[133] The more common pattern is one of selective application by States of those duties and obligations of membership which seem to conflict with national interests. Given the wide range of duties and obligations which flow from Community membership such practice, if adopted, would be lethal to the Communities. Equally common is the failure of international organizations to adopt sanctions against breach and – on those occasions when measures are adopted – to enforce them effectively. The essence of what for convenience I call the "all-or-nothing effect" in the European Communities is that whereas Member States retain the ultimate political option of withdrawing from the Community and thereby disengaging from their obligations of membership (an option which the process of economic and political enmeshment has made increasingly theoretical), they are – as long as they opt for membership – largely unable to practise selective application of Community obligations. There still remain, for reasons which will be explained below, certain lacunae in the full realization of the "effect", but it has certainly reached a stage where it can be stated as a fundamental distinguishing mark of supranationalism in both its facets.

We have already noted the central role of the Court of Justice in the evolution of supranationalism.[134] But the existence of a court as part of the institutional framework of an international organization is not unique and

cannot, as such, explain the "all-or-nothing effect", nor can the existence of a compulsory jurisdiction *per se* be sufficient explanation since – as recently exemplified[135] – submission to the compulsory jurisdiction and subsequent obedience to an award are among the obligations which in the current state of international law can often be flouted with impunity. Rather it is the entire *system* of judicial review involving both national courts and the European Court, which transnationalizes mechanisms hitherto used only in the context of municipal judicial review,[136] which produces the "effect".

## 2. *The functional division of adjudicatory tasks and judicial review*[137]

The Community features a double-limbed system of judicial review which operates on two levels. Two sets of legislative acts and administrative measures are subject to judicial review: (a) (the first limb) those of the Community legislative and administrative institutions (principally Council and Commission) which are reviewable for conformity with the provisions and principles of the Treaties and with an emerging unwritten higher law based on the constitutional traditions of all Member States as well as international treaties such as the European Convention on Human Rights;[138] and (b) (the second limb) acts of the Member States which are reviewable, in accordance with the principle of supremacy, for conformity with Community law itself. Needless to say, in the context of compliance of Member States with Community obligations, effective review of the latter set is the crucial issue.[139]

Judicial review may take place at the exclusive level of the Community Court. As regards the first – less critical – limb, the organs of the Community and the Member States, as well as individuals, may, in accordance with the Treaty,[140] challenge Community acts and measures directly before the European Court. As may be expected the rules of standing for individuals are quite narrowly defined and the Court has added to this narrowness by interpreting them rather strictly.[141]

As regards the second – critical – limb the Commission and Member States may, in accordance with the Treaty,[142] bring an action against a Member State for failure to fulfil its obligations under the Treaty. Failure to fulfil an obligation may take the form of inaction in implementing a Community obligation or enacting a national measure contrary to Community obligations. Although, as indicated above, not unique, the very existence of a non-optional judicial forum for adjudication of these types of disputes sets the Community above most international organizations. At the same time the "intergovernmental" character of this process and the consequent limitations on its efficacy are clear enough. Four weaknesses are particularly glaring.

In the first place, the decision of the Commission and/or Member States to bring an action against an alleged violation by another Member State will often be influenced by political considerations; the Commission might not wish to risk a political crisis which may be precipitated by a Court decision on a sensitive issue.[143] Secondly, effective supervision will depend on the ability of the Commission to monitor the implementation of Community law. Given the vast range of Community measures this becomes an impossible task.[144] Thirdly, the type of action which is likely to be brought will relate largely to abject Member State failure to *implement* a national measure required by Community law or to a national measure which is in clear violation of Community law. It will be far less suited to review of the ordinary application and enforcement by Member States of Community obligations especially as they affect individuals.[145] That type of violation becomes normally transparent only through cases and controversies affecting individuals. Even if alleged violations were brought to the attention of the Commission, it is unrealistic to expect them to take up all but the most flagrant violations. Finally, given the intergovernmental character of this process, a Member State found to have failed to fulfil an obligation may simply disregard the judgment against it.[146]

These weaknesses are to an extent remedied by the review of both limbs at the national level, a process possible through the functional division of judicial tasks between the European Court of Justice and national courts and which essentially produces the "all-or-nothing effect" It is hardly worth mentioning that of all Treaty provisions the single most important Treaty Article is 177.[147] This multi-functional Article provides *inter alia* that when a question concerning either the interpretation of the Treaty or the validity and interpretation of acts of the institutions of the Community is raised before national courts, the latter may (and in the case of courts of final instance, must) refer the issue for a preliminary ruling of the European Court of Justice. Once this ruling is made it will be remitted back to the national court which will give, on the basis of the ruling, the decision in the case pending before it. The national courts and the European Court of Justice are thus integrated into a unitary system of judicial decision-making. The two limbs of judicial review exist on this level as well. A reference to the Court on the *validity* of acts of institutions is clearly a mode for judicial review of Community acts at the instance of individuals. One will note that the question of *locus standi* from the point of view of the European Court does not arise. Thus it may even be possible for an individual to be denied standing in a direct challenge before the European Court but have the act reviewed if he has standing in accordance with national procedural law. The individual will be able to challenge the Community act in the national courts (Community law being, of course, part of the "law of the land"), whereupon it will be

remitted to the European Court of Justice for an interpretation on validity and returned back to the national court for pronouncement.[148] Taking then judicial review of Community measures as a whole, the trend, in respect of individual challenges, is one of a restrictive attitude to actions brought directly before the European Court with a shift to national courts as the forum for adjudication, using, when necessary, the preliminary reference for an interpretation or check of validity, of a Community measure.[149]

Turning to the second limb concerning the judicial review of national measures for conformity with Community law, the European Court has made astute use of that part of Article 177 which provides for references on the "interpretation" of Community law. On its face the purpose of the procedure is to guarantee uniform interpretation of Community law in all Member States. However, often the factual situation in which Article 177 is employed is when a litigant pleads in the national court that a rule or measure of national law or an administrative practice, should not be applied as it is in contradiction with Community law. On remission to the European Court it renders its interpretation of Community law within the factual context of the case before it. Theoretically, a division exists in the adjudicatory tasks of the two courts: the European Court states the law and the national court applies it – using of course the principle of supremacy where necessary – to the case in hand. Thus the traditional formula of the Court is to state that:

> ... The Court of Justice ruling under Article 177 of the EEC Treaty does not hold that a given national [law] is incompatible with Community law... Within the context of the judicial cooperation established by the provision it is for the national courts, applying the fundamental rule that Community law takes precedence, to uphold the right, of the subjects based, under the Treaty itself on the direct effect of [the provisions concerned] when disputes are brought before them by those concerned.[150]

But as usefully concluded in a study on the role of the European Court in judicial review:

> It is no secret, however, that in practice, when making preliminary rulings the Court has often transgressed the theoretical borderline... it provides the national judge with an answer in which questions of law and of fact are sufficiently interwoven as to leave the national judge with only little discretion and flexibility in making his final decision.[151]

What is important – indeed crucial – is the fact that it *is* the national court acting in tandem with the European Court which gives the formal final decision on the compatibility of the national measure with Community law. The main result of this procedure is the binding effect and enforcement value which such a decision will have on a Member State – coming from its own courts – as opposed to a similar decision handed

down from Luxembourg by the European Court of Justice wearing its intergovernmental hat. This then is a procedural dimension of the constitutionalization of the Treaties and a European confirmation of Mr Justice Douglas' maxim that "... it is procedure that marks much of the difference between rule by law and rule by fiat".[152] It is also one of the clearest distinguishing marks of the supranational system. The quest for an effective law of nations in the traditional international legal order has been characterized by the creation of a succession of international courts, tribunals, arbitration bodies and other judicial and quasi-judicial fora. With a few exceptions these bodies have all been victims of the inherent weakness of international judicial bodies by comparison to their national counterparts. International jurisprudence, with all the attention it receives from scholars, has remained on the periphery of international law and international relations. By contrast, the supranational system – in a synthesis of international law and constitutional law – puts the inherently stronger *national* system in the service of the transnational order.

The above analysis helps to express the current limitations of the evolution of the "all-or-nothing effect" in the Community. If all issues involving alleged violations of Member States' Community obligations became matters involving private parties triable before national courts the "all-or-nothing effect" could be said to be complete. Inevitably, however, there are certain matters which concern directly Member States only and which can not realistically become case-and-controversy issues. Violation of these would still remain a matter for "Community level" intergovernmental judicial review. In addition, even in those situations where an individual action – a case and controversy – could, by virtue of the subject matter, take place, there may be an array of additional barriers to overcome. These may be connected with ignorance by the individual of his "higher" Community rights, the barriers of expense and time of litigation and the low stakes which the individual may have in vindicating a right despite its wider Community importance. (In *Costa* v *ENEL* the actual controversy concerned a sum of £1).

Finally, the use of Article 177 as a method for judicial review of Member State measures depends on the full acceptance by the national courts of the doctrine of supremacy and on a willingness to utilize 177 to its full potential. We have noted that both in Britain and in the administrative judicial branch in France the first issue is not clearly resolved and as regards the latter there are mixed trends. Nevertheless, even if incomplete, the existing "all-or-nothing effect" remains a singular expression of supranationalism distinguishing the Community from most other international organizations. Its preservation and consolidation must surely be one of the important challenges of the future.

## Notes

[1] The Treaty establishing the European Coal and Steel Community (hereinafter: Treaty of Paris) which launched the Community experience was signed in Paris on 18 April 1951. Treaties establishing the European Economic Community (hereinafter: Treaty of Rome) and the European Atomic Energy Community (hereinafter: Euratom) were signed in Rome on 25 March 1957. The Schuman Declaration of 9 May 1950 may be regarded as the starting point: See 5/6 *European Community* 3 (1980) and see also *Bulletin of the European Communities*, 13-1980, point 1.2.1. But see note 36 *infra*.
Given the measure of institutional integration among the three Communities I shall refer to the combined structure as the Community.

[2] The treatment here will be confined solely to the Community experience although other regional organizations such as the Council of Europe have undoubtedly played a role in European integration and display certain supranational features. See, for example, Drzemczewski, "The *sui generis* nature of the European Convention on Human Rights", 29, *ICLO*, 54, 1980.

[3] Greece acceded formally on 1 January 1981. The Treaty of Accession was signed in Athens on 28 May 1979. For details of the conditions of accession and the main provisions in that Treaty, see *Bulletin of the European Communities*, 12-1979, points 1.1.1-1.1.19.
Spain and Portugal are candidates for accession within this decade.

[4] For a useful analysis of the possible impact of, and problems created by, accession, see D. Marquand, *Parliament for Europe*, Jonathan Cape, London, 1979, especially at pp. 30-33.

[5] One should not exaggerate the potential for immediate change in the institutional balance which the new directly elected Parliament may have. For a restrained and cautions analysis emphasizing "...the limits to the influence upon public policy of a directly elected European parliament", see "The policy implications of direct elections" (various authors), 17, *Journal of Common Market Studies*, 1979, at pp. 281-349.
At the same time the Parliament has already exhibited its constitutional aggressiveness and legal astuteness by such acts as rejecting the 1980 budget *in toto* and, in a more subtle manner, voting an increased supplementary budget in 1980 deliberately so as to use unspent surpluses in 1981 [see 16 *European parliament*, column 1 (1980)]. This last move has precipitated a political crisis whereby Belgium, France and Germany have refused to make their full Community budgetary contributions. The Commission has commenced legal proceedings against the recalcitrant States. In addition, the Parliament, astutely using a provision in the Statute of the Court of Justice (Article 37) has managed to intervene in Court proceedings in which a Council regulation (Regulation No 1293/79) was contested on the grounds that Parliament was not consulted, contrary to treaty requirements relating to that class of measure. The Court upheld the parliamentary contention. See Joined Cases 138 and 139/79 *Maizena Gesellschaft* v *Council*; *Roquette Frères* v *Council*. Decision of 29 October 1980 (not yet reported).

[6] "For years the Community has been described as being in crisis. But when crises exist permanently, merely changing their immediate causes, it should be asked if they really are crises, that is to say exceptional conflict situations. It is rather more likely that the conlicts the Community has so far experienced are significant of tensions inherent in the integration process itself". Everling, "Possibilities and limits of European integration", 18, *Journal of Common Market Studies*, 1980, at p. 217.
Whilst I accept Everling's analysis of the Community as a system of crises my argument is not that the present problems and challenges are inherently more difficult than previous ones, but rather that conditions for solution – such as the present economic climate – have worsened.

[7] Of course, when assessing the Labour Party decision, allowance must be made for the

traditional licence of opposition parties out of government. For the Debré-Foyer Bill see *Proposition de loi portant rétablissement de la souveraneté de la République en matière d'énergie nucléaire*, No 917, Assemblée nationale, 2e session extraordinaire de 1978-79.

[8] Case 232/78 *Commission* v *French Republic* (1979) ECR 2729 in which the judgment was given, and Joined Cases 24/80 R and 97/80 March 1980 *Commission* v *French Republic* in which non-compliance was established (even though the Commission failed in this latter case to get interim measures against France). See "Editorial comments, The mutton and lamb story: Isolated incident or the beginning of a new era?", 17, *CML Rev.*, 1980, at p. 311.
A case of non-complicance occurred when Italy failed to follow judgment against it in Case 7/68 *Commission* v *Italian Republic* (1968) ECR 4 and thus had to be prosecuted again, Case 48/71 *Commission* v *Italian Republic* (1972) ECR 532.
This, however, was an instance of dilatoriness rather than outright defiance as in the Mutton Case. Even more dangerous has been French judicial defiance by its Conseil d'État which, apart from rejecting the principle of supremacy (to be discussed *infra*), has rejected positive principles of Community law as decided by the Court of Justice. See *Cohn-Bendit case* (December 1978 D 79 J 155). This defiance is confined to the administrative branch of the judiciary.

[9] Although the term supranationalism is used commonly in the literature, within the framework of the Treaties establishing the Community it is only mentioned in the Treaty of Paris (Aticle 9). The term, unjustifiably, became associated too much with extreme integrationalism and was dropped in the Treaty of Rome and the Euratom Treaty.

[10] Robertson, "Legal problems of European integration", 91, RDC, 105, at p. 143 (1957).

[11] A. Shonfield, *Europe: Journey to an unknown destination*, Allen Lane, London, 1972, at p. 17.

[12] In this context I am using the term "federal" in its widest, most fundamental sense of sharing in governance over activities. Elazar usefully records the origins of the term "...first in the biblical hebrew term *brit*, then the latin *foedus* (literally 'covenant'), from which the modern 'federal' is derived... Elaborated by the Calvinists in their federal theology, the concept formed the basis for far more than a form of political organization... The original use of the term deals with contractual linkages that involve power sharing – among individuals, among groups, among States. This usage is more appropriate than the definition of modern federations, which represents only one aspect of the federal idea and one application of the federal principle".
D. Elazar (ed.), *Self rule/Shared rule*, Turtledove, Ramat Gan, 1979, at p. 3. This overview of the federal principle is particularly important since although the Community is, in this wide sense, a "federal" entity, it decidedly does not conform to the traditional notion of having (or aspiring to have) a strong important centre and periphery linked thereto.

[13] The Community is new in the post-World War II period. The nineteenth-century German Zollverein (Keeton, "The Zollverein and the Common Market", in Keeton and Schwarzenberger (eds.), *English law and the Common Market*, Stevens and Sons, London, 1963, 1), and the Danube Commission (Smith, Danube, *4 yearbook of world affairs* (1950), at p. 191) were two antecedents in earlier days. The term supranationalism also predates the Community. Schermers cites Einstein as wiriting to Freud in 1932 and stating "At present we are far from possessing any supranational organization". H. G. Schermers, *Internation institutional law*, Vol. 1, Sijthoff, Leiden, 1972, at p. 20. It is, of course, possible to adopt an a priori definition but this theoretical approach on a choice of criteria which will necessarily be subjective. Schermers prefers this approach adopting a useful list of such criteria but even he is then pushed to conclude that since to be "completely supranational, an international organization should fulfil all [(these criteria)]... no such supranational organization exist», op. cit., p. 21. I have preferred a more inductive approch relying on

the experience of the Community itself even if the results is not as theoretically satisfying as the a priori method. See also, G. Mally, *The European Community in perspective*, Lexington Books, Lexington, 1973, especially at p. 26.

[14] This is perhaps my bias as common law lawyer: "The pursuit of definitions has never appealed much to lawyers because they are aware that the concepts they employ have been rough-newn by history and stoutly resist philosophical formulation". Pollock, "The distinguishing mark of crime", 22, M.L.R., 495 (1959):

[15] P. Hay, *Federalism and supranational organizations*, University of Illinois Press, Urbana and London, 1966. In its historical synthesis parts Hay's study offers the most exhaustive treatment of different studies of supranationalism especially in its legal and institutional aspects.
Moreover, the analytical parts of the study have retained their value despite the passage of years. Even today the book repays careful study. I have relied on Hay for the brief survey of different treatments in the present retrospective analysis.

[16] Op cit., Chapter 2, and appendices, pp. 77-78.

[17] See for example, Nørgaard, *The position of the individual in international law*, Munksgaard, Copenhagen, 1962, (International Law Approach) and Kohnstamm, *The European Coal and Steel Community* 90 *RDC* 1 (1956 II) and comments thereon in Hay, op. cit.

[18] "...an unsatisfying shrug", Hay, "Federal jurisdiction of the Common Market Court", 12 *Am. J. Com. L.*, 1963, p. 39. Cf. Hay, op. cit., p. 44 and note 106.

[19] Thus in a leading current treatment Schermers has no hesitation in including the Community in his general treatise on international institutional law but he is careful to distinguish between supranational and intergovernmental organizations. See Schermers, note 13 *supra* at pp. 19-24.

[20] e.g. Robertson, note 10 *supra*, p. 145.

[21] In their comprehensive collection of texts, cases and readings, Stein, Hay and Waelbroeck suggest six phases (The phase of enthusiasm 1945-49; Towards integration by sectors 1950-55; Relaunching 1955-58; A split – A bridge – political "relaunching"? 1958-63; Crisis-consolidation 1963-68; Enlargement 1969-??). E. Stein, P. Hay, M. Waelbroeck, *European Community law and institutions in perspective*, Bobbs-Merrill, Indianapolis, New York, 1956, pp. 10-13. Greilsammer, suggesting that "there is virtually no process that can be delimited in time as well as the process of European integration" opts for four periods 1946-50; 1950-58; 1958-69; 1969-??. Greilsammer, "Theorizing European integration in its four periods", 2 *The Jerusalem Journal of International Relations* 129 (1976) (for the evaluation of the different periods se text to notes 25-26 *infra*). Dahrendorf suggests three less well-defined phases: The Founding Fathers – Monnet *et al.* (approx. 1950s); Founding sons – Hallstein *et al.* (approx. 1960s), and the present generation. R. Dahrendorf, *A third Europe?* EUI, Florence, 1979 – Jean Monnet lecture.
Pryce also suggests six phases although different from Stein *et al.*: 1950-51; 1952-54; 1955-57 (Relance); 1958-62 (New Communities in (action); 1963-69 (Conflict, crisis and stagnation); 1969-72 (Second relance). R. Pryce, *The politics of the European Community*, Butterworths, London, 1973, pp. 1-27.

[22] Stagnation of supranationalism, howeer, does not imply stagation of substantive integration.

[23] Greilsammer, op. cit.

[24] At p. 141 (emphasis added).

[25] The major decision on direct effect was given on 5 February 1963. On supremacy on 15 July 1964. For more detailed discussion, see text to note 41 *infra*.

[26] The period of De Gaulle in which UK accession was rejected (on a French "veto") and in which the Luxembourg crisis occurred created a general overhaul of theories of integration. See, for example, Haas, "The uniting of Europe and the uniting of Latin

America", 5 *Journal of Common market Studies* 315 (1967), at pp. 325-331.
An extremely pessimistic assessment in the mid-1960s – with statements such as "Supranational structures may not survive into 1966" – is that of Heathcote, "The crisis of European supranationality", 5 *Journal of Common market Studies* 140 (1966). The analysis, strongly influenced by the political crises of that period, is instructive in illustrating the cleavage. Heathcote, at p. 141, adopts an a priori definition of supranationality according to which "a supranational organization is one which (a) bypasses the nation-State's authority and deals directly with the citizens; which (b) takes over some functions traditionally exercised by the nation-State; and (c) is in the position to originate decisions not only on behalf of the State but despite it". It is interesting that there is almost exclusive concentration on the decision-making actors and processes and only oblique – if at all – reference to the validity and status of the decisions adopted *vis-à-vis* national measures. The latter are the traditional preoccupations of the lawyer. And yet without this latter type of validity the power to originate measures despite Member State opposition would have precarious value if these measures could subsequently be overturned by a member State change of mind; Heathcote's first criterion – the authority to deal directly with the individual – introduced already by the Treaty of Paris (a power which, incidentally, remained largely intact during the 1960s), has been overtaken by developments in the 1960s, by decisions on self-executing measures and supremacy which are more revolutionary, have greater impact and could far better serve as distinguishing criteria for supranational organizations.
Puchala, although dealing with the wider concept of international integration captures with his "blind men and elephant" metaphor neatly, if somewhat acidly, the problem of the disciplinary cleavage: "Each blind man... [(touching)] a different part of the large animal, and each [(concluding)] that the elephant [(international integration)] had the appearance of the part he touched". Puchala, "Of blind men, elephants and international integration", 10 *Journal of Common Market Studies* 267 (1972) at p. 267. His own sophisticated "concordance system" strangely pays little attention to the constitutional developments which could have been regarded as important supportive elements, at pp. 277-284, but see at pp. 269-271.

[27] Note 26 *supra*, and Greilsammer, note 21 *supra* at pp. 142-146.

[28] "Constitutionalization" implies a combined and circular process by which the Treaties were interpreted by techniques associated with constitutional documents rather than multi-partite treaties and in which the Treaties both as cause and effect assumed the "higher law" attributes of a constitution. For an interesting discussion see *Proceedings of the 72nd Annual Meeting of the American Society of International Law*, at pp. 166-197 (1978).
The German Federal Constitutional Court has actually said that "The European Economic Community Treaty is, as it were, the constitution of this Community"; Federal Constitutional Court, First Chamber, Decision of 18 October 1967; ((1967)) AWD 477-78; ((1968)) *Europarecht* 134-37 cited by Stein in *Proceedings*, op. cit., p. 168.

[29] See, for example, Pescatore, note 32 *infra*.

[30] Hay, note 15 *supra*, p. 69.

[31] Hay, op. cit., p. 31 ff.

[32] Cf. P. Pescatore, *The law of integration*, Sijthoff, Leiden, 1974. In his excellent study Judge Pescatore tends to play down the institutional crises (e.g. pp. 11-19) such as the Luxembourg accord. Consequently his treatment gives a general impression of continuing progressive evolution.

[33] I prefer "decisional" to "institutional" since the former conveys the need to look at the actual processes and not merely at formal functions. Far more sophisticated tools and frameworks have been offered for the analysis of federal models in general and the

Community model in particular. Elazar's series of matrixes is a recent most valuable contribution as regards the former (see Elazar, "The role of federalism in political integration" in D. J. Elazar (ed.), *Federalism and political integration*, Turtledove, Ramat Gan, 1979, 13. See also W. H. Riker, *Federalism: Origin, operation, significance*, Little, Brown, Boston, 1964. Lindberg's model has become something of a classic as regards the latter (see Lindberg, "The European Community as a political system: Notes toward the construction of a model", 5 *Journal of Common Market Studies* 359 (1967) and L. Lindberg and S. Scheingold, *Europe's would-be policy: Patterns of change in the European Community*, Prentice Hall, New Jersey, 1970. The limited framework here is probably sufficient for our purpose since it concentrates on supranationalism in its instrumental facet and not on the uses to which it has been put in the evolvement of substantive policies and the evaluation thereof. The limited framework will also enable me, in the confines of this essay, to flesh it out so that it does not remain too abstract.

[34] A fully fledged decisional analysis would have to take separately each single policy, determine the factors, forces and actions relevant thereto and attempt to trace the decision-making process. Puchala, note 26 *supra* at p. 278, has constructed such a model as regards the agricultural sector. Inevitably, this cannot be done here and I have to content myself with a general Community analysis even if at a great sacrifice of sophistication. It is submitted, however, that the general Community analysis remains relevant to individual sectors. Naturally I do not claim that this framework can achieve precise measurement. It fails, thus, one of Deutsch's crucial tests for 'theoretically powerful' models (see K. Deutsch, *The nerves of government*, The free Press of Glencoe, New York, 1963, Chopur 1). Still, as providing a rough measuring instrument enabling at least the indications of trends, it may be considered adequate for this retrospective analysis. Detailed frameworks such as Puchala's carry the danger of remaining too theoretical and incapable of practical applciation. Thus, in H. Wallace, N. Wallace and Webb (eds), *Policy-making in the European Community* (John Wiley and Sons, London, 1976), which analyses policy-making in specific fields, the authors, including Puchala himself, had to adopt less detailed frameworks.

[35] Lindberg, in his 'scale of decision locus', note 33 *supra*, at pp. 356-357, offers a more comprehensive breakdown along a '... continuum ranging from decisions taken entirely or almost entirely in the Community system' to 'deicisions taken entirely by the national systems individually'. The full range consists of:

1. Decisions... taken entirely in the [EC] system.
2. Decisions... taken almost entirely in the [EC] system.
3. Decisions... taken predominantly in the European Community system, but the nation-States play a significant role in decision-making.
4. Decisions are taken about equally in the European Community system and the nation-States.
5. Decisions are taken predominantly by the nation-States, but the European Community system plays a significant role in decision-making.
6. Decisions are taken almost entirely by the nation-States.
7. Decision are taken entirely by the nation-States individually.

This model less useful for us, since its main purpose is to determine in relation to a list of *substantive functions* which political systems fulfil, the degree to which the Community is 'substituting' the Member States. In Legal terms its purpose would be to delineate substantive Community jurisdiction and competence. It does not focus on the decision-making process itself and thus its utility here may be questioned. Since, if there is a measure of truth in the assessment "...that the Council [of Ministers] is in fact no longer a Community institution, but only a sort of clearing house for national interests, which by using the principle of unanimity prevents any further progress of the European Communi-

ty" (Aigner, Member of the European Parliament, *Debates of the European Parliament*, 10.7.1980, p. 290 (English version)), then in terms of the instrumental means of supranationalism, the fact that the locus of decision falls within Lindberg's first category, becomes less meaningful. Lindberg's "scale of peripheralization-centralization" (drawing on Riker) and of systems and subsystems goes some way towards this decisional analysis but is, again, too detailed to be of use in a limited survey.

[36] Articles 14, 15 ESC. Interestingly, the Schuman Declaration merely states: "Par la mise en commun de productions de base et l'institution d'une Haute Autorité nouvelle, dont les décisions lieront la France, l'Allemagne...", indicating decisions binding *on* States and not *in* States. The formal supranational leap was effected by the actual Treaty framers who gave the High Authority power to adopt measures directly effective in the legal order of the Member States.

[37] Decisions binding on members may be taken by UN organs. See, for example, *Charter of UN*, chapter VII. For commentary, see Y. Dinstein, *International law*, Vol. 5, pp. 53-57, Schocken, Tel Aviv, 1979.

[38] Robertson, note 10 *supra*.

[39] See note 13 *supra*. Religious law – especially where given exclusive jurisdiction in, say, family matters – may be regarded as supranational in this sense. The Catholic Church and Jewish Rabbinate could therefore be regarded in old and modern times as being supranational.

[40] The literature on the doctrine is immense. For a lucid up-to-date statement see, for example, D. Wyatt and A. Dashwood, *The substantive law of the EEC*, Chapter 3, Sweet and Maxwell, London, 1980.
For wider studies, see Waelbroiek, "Effets internes des obligations imposées à l'État", in *Miscellanea W. J. Ganshof Van Der Meersch*, Volume 2, Bruylant, Brussels, 1972, 573. Bebr, "Directly applicable provions of Community law: The development of a Community concept", 19 *ICLO* 257 (1970).

[41] Case 26*/62 *Van Gend en Loos* v *Nederlandse Administraitie der Belastingen* [1963] ECR 1.

[42] The main operative part of the judgment is so well known as to render citation almost superfluous. For the benefit of non-Europeans, the following are the key elements in the judgment:
"The objective of the EEC Treaty, which is to establish a common market, the functioning of which is of direct concern to interested parties in the Community, implies that this Treaty is more than an agreement which merely creates mutual obligations between the Contracting States. This view is confirmed by the preamble to the Treaty which refers not only to governments but to peoples. It is also confirmed more specifically by the establishment of institutions endowed with sovereign rights, the exercise of which affects Member States and also their citizens. Furthermore, it must be noted that the nationals of the States brought together in the Community are called upon to cooperate in the functioning of this Community through the intermediary of the European Parliament and the Economic and Social Committee... This confirms that the States have acknowledged that Community law has an authority which can be invoked by their nationals before those courts and tribunals... The conclusion to be drawn from this is that the Community constitutes a new legal order of international law for the benefit of which the States have limited their sovereign rights, albeit within limited fields, and the subjects of which comprise not only Member States but also their nationals";
at p. 12.

[43] In Case 127/73 *Belgische radio en Televisie* v *SABAM* [1974] ECR 51, the European Court held that Articles 85 and 86 EEC were capable of bestowing rights and duties on individuals *inter se*. It should be noted, however, that these Treaty articles themselves

involved actions of individuals. Cf. Case 13/61 *Bosch* v de Geus [1962] ECR 45 in which this development is already anticipated. (See Wyatt and Dashwood, note 40 *supra* at pp. 29-30). The doctrine was evolved further in a subsequent case which was concerned with the general principle embodied in Article 7 EEC (non-discrimination on grounds of nationality) and which, unlike SABAM, did not necessarily involve individuals. Even the Commission – usually very integrationalist-minded – doubted whether this Treaty principle should be given "horizontal effects". The Court took the radical position and held that the Treaty could indeed bestow rights and duties on individuals *inter se*. Case 36/74 *Walrave and Koch* v *Association Union Cycliste Internationale* [1974] ECR 1405. See also Case 43/75 *Defrenne* v *Sabena* [1976] ECR 455.

[44] It is not proposed to discus here the well known distinction between direct applicability and direct effect. (See, for example, Winter, "Direct applicability and direct effect, Two distinct and different concepts in Community law, 9 *CMI Rev.* 425 (1972)). The extension of direct effect to directives was remarkable. Whereas regulations by virtue of Article 189 EEC are directly applicable and thus inevitably, if self-executing, produce – by analogy to the reasoning of the Court in relation to Treaty provisions – automatic direct effect, directives are only binding as to the result but leave to the national authorities the choice of form and method. It may then have been thought that they could not produce direct effect. See *Joseph Aim and Société SPAD* v *L'Administration des douanes* [1972] *CMLR* 901. The Court of Justice, in a step-by-step approach, has applied the doctrine even though subject to possible different structural conditions (note 155 *infa*) to directives as well. Cases signalling this evolution are: Case 9/70 *Franz Grad* v *Finanzamt Traunstein* [1970] ECR 825 (direct effect of a time limit in a directive – "vertical" effect); Case 41/74 *Van Duyn* v *Home Office* [1974] ECR 1337 (direct effect of a substantive provision of a directive but one which elaborated a substantive right bestowed by the Treaty – "vertical" effect); Case 51/76 *Verbond van Nederlandese Ondernemngen* v *Inspctor der Invoerrechten en Accijnzen* [1977] ECR 113 (direct effect of substantive provision of directive concerning an obligation not directly bestowed by the Treaty *and* against which a national implementing measure was reviewed – "vertical" effect). Cf. Case 21/78 *Delkvist* [1978] ECR 2327). On the possible extension of "horizontal" direct effect to directives see Easson, "Can directives impose obligations on individuals", 4 *EL Rev.* 67 (1979). But see now case 148/78 *Rattu* [1979] ECR 1629 (Advocate General submissions) and Usher, "The direct effect of directives", *EL Rev.* 268 (1979); see also Timmermans, note 155 *infra*.

[45] Case 6/64 *Costa* v *ENEL* [1964] ECP 585. With the same reservations expressed in note 42 *supra* the following are the main operative elements in the judgment:
"By contrast with ordinary international treaties, the EEC treaty has created its own legal system which, on the entry into force of the Treaty, became an integral part of the legal system of the Member States and which their courts are bound to apply. By creating a Community of unlimited duration, having its own institutions, its own personality, its own legal capacity and capacity of representation on the international plane and, more particularly, real powers stemming from a limitation of sovereignty or a transfer of powers from the States to the Community, the Member States have limited their sovereign rights, albeit within limited fields, and have thus created a body of law which binds both their nationals and themselvels. The integration into the laws of each Member State of provisions which derive from the Community, and more generally, the terms and the spirit of the Treaty, make it impossible for the States, as a corollary, to accord precedence to a unilateral and subsequent measure over a legal system accepted by them on a basis of reciprocity. Such a measure cannot therefore be inconsistent with that legal system. The executive force of Community law cannot vary from one State to another in deference to subsequent domestic laws without jeopardizing the attainment of the objectives of the

Treaty set out in Article 5 (2) and giving rise to the discrimination prohibited by Article 7";

at pp. 593-594.

[46] Another important instance of constitutional interpretation occurred in Case 38/69 *Commission* v *Italian Republic* [1970] ECR 56 in relation to Community "secondary" legislation. The Italian Government argued that a certain internal Community measure (pursuant to Article 235 EEC and accelerating the realization of the common market) had a *contract* basis as between the Member States and constituted an international agreement to which even reservations could be made. The Court gave short shrift to the argument upholding the *institutional* rather than contractual nature of Community measures.

[47] Here, of course, we have one of the most intractable problems of Community law. The Treaty of Rome is in many of it provisions fairly general lending itself to expansive teleological interpretation by the Court. This coupled with certain "elastic" clauses (e.g. Article 235 EEC) gives a wide measure of latitude to the policy-making organs to extend the boundaries of Community competence. Often this meets with national resistance. Cf. Close, "Harmonization of laws: Use or abuse of the powers under the EEC Treaty?, 3 *EL Rev.* 461 (1978).

[48] Case 14/68 *Walt Wilhelm and Others* v *Bundeskartellamt* [1969] ECR 1.

[49] Case 106/77 *Italian Finance Administration* v *Simmenthal* [1978] ECR 629.

[50] Recital 3 of judgment.

[51] Recitals 7-9 of judgment.

[52] Se P. Barile (ed.), *Il primato del diritto comunitario e i giudici italiani*, Franco Angeli, Milano, 1978.

[53] The case involved, naturally, a combination of direct effect and supremacy insues. In relation to supremacy the Court stated in recitals 17-23 that:

"In accordance with the principle of precedence of Community law, the relationship between provisions of the Treaty and directly applicable measures of the institutions on the one hand and the national law of the Member States on the other is such that those provisions and measures not only by their entry into force render automatically inapplicable any conflicting provision of current national law but – in so far as they are an integral part of, and take precedence in, the legal order applicable in the territory of each of the Member States – also preclude the valid adoption of new national legislative measures to the extent to which they would be incompatible with Community provisions... Every national court must, in a case within its jurisdiction, apply Community law in its entirety and protect rights which the latter confers on individuals and must accordingly set aside any provisions of national law which may conflict with it, whether prior or subsequent to the Community rule. *Accordingly any provision of a national legal system and any legislative, administrative or judicial practice which might impair the effectiveness of Community law by withholding from the national court having jurisdiction to apply such law the power to do everything necessary at the moment of its application to set aside national legislative provision which might prevent Community rules from having full force and effect are incompatible with those requirements which are the very essence of Community law.* This would be the case in the event of a conflict between a provision of a Community law and a subsequent national law if the solution of the conflict were to be reserved for an authority with a discretion of its own, other than the court called upon to apply Community law, *even if such an impediment to the full effectiveness of Community law were only temporary*"

(emphasis added).

The internal Italian provision decided upon by the Italian Constitutional Court whereby any question of conflict between Community law and national law was a constitutional issue, resolution of which must be decided by that Constitutional Court itself by an internal reference from the lower court, can be criticized in terms of the general efficiency

of the Italian legal system. It resembles the Amparo institution which led to the "degeneration" of an important constitutional mechanism in the Mexican legal order. See M. Cappelletti, *Judicial review in the contemporary world*, Bobbs-Merrill, Indianapolis, 1971, at pp. 20—21.

From the point of view of European law, the European Court's decision contains a possible ambiguity since, as seen, its ruling is applicable to a "... national court *having jurisdiction to apply such law...*". It could possibly be argued that in cases of conflict the lower national court, by virtue of the decision of the Italian Constitutional Court, has no such jurisdiction. Since the jurisdictional competence of courts is usually a matter for the national legal order, the Court of Justice could not interfere in this matter any more than if certain lower courts were denied jurisdiction over matters involving a large sum of money even if concerned with Community law.

[54] See generally, Behr, "How supreme is Community law in the national courts?", 11 *CML Rev*: 479 (1974) 3.

In the Netherlands and Luxembourg, having a monist system which acknowledges the supremacy of treaty law, acceptance was not difficult. (See Articles 63 and 65, 66, 67 of the Dutch Constitution.) Application of the European Court's decision in, say, *Van Gend en Loos* was in fact non-problematic. As regards Luxembourg see Pescatore, "Prééminence des traités sur la loi interne selon la jurisprudence luxembourgeoise" [1953] *Journal des Tribunaux* 455.

In Belgium, the situation was constitutionally ambiguous until the landmark decision of the Belgian Cour de Cassation in the *Le Ski case* [1972] *CMLR* 373 in which the Belgian Supreme Court adopted the most full-blooded version of supremacy as required by the European Court. The question of a Community provision coming into direct conflict with norms of the Belgian (and French) Constitution arose in a recent decision – Case 149/79 *Commission* v *Kingdom of Belgium*, Decision of 17 December 1980 (not yet reported). The Court remitted the case back to the parties for further clarification before final resolution. The Belgian Government in its pleading did not deny that Community rules override national rules but suggested that in interpreting the meaning of a term in the Treaty (in that case "public service") the Court should use an approximation of the constitutional law of the Member States as an interpretative aid.

[55] This was the case in Germany and Italy. See *German handelsgesellschaft case* [1974] 2 *CMLR* 551 (Decisions of 29 May 1974 BVerfG 37; 271); and *Italian Frontini case* [1974] 2 *CMLR* 386. For a useful discussion on the implications of this case see H. G. Schermers, *Judicial protection in the European Communities*, Kluwer, Deventer, 1979, pp. 92-97. Note that the German Federal Constitutional Coourt was concerned with constitutional safeguards as regards legislation; the supremacy challenge was only indirect. Otherwise, the German Federal Constitutional Court has fully accepted the supremacy of Community law even over subsequent national law – see *German Lütticke case*, Bundesverfassungsgericht Decision of 9 June 1971, [1971] AWD, 418-420 (BVerfG 31; 145). The German federal Constitutional Court also rejected the possibility of *Verfassungsbeschwerde* (constitutional complaints) as against acts of the Community authorities limiting this type of recourse to action by German public authorities. See *german Constitutional Rights case*, Bundesverfassungsgericht decision of 18 October 1967, [1967] AWD 477 (BVerfG 22; 293); and note in 5 *CML Rev*. 483 (1967-68).

[56] However, unless the German federal Constitutional Court modifies its own position (cf. now BVerfG decision of 25 July 1979, 15 *Europarecht* 68 (1980)) the conflict cannot be fully resolved until the Community has a written bill of rights which corresponds to the guarantees of the German basic law. Another condition which the German Federal Constitutional Court imposed, which has been only partially fulfilled is the evidence of a democratically legitimated parliament directly elected by general suffrage (which now

exists) *which possesses legislative powers and to which the Community organs empowered to legislate are fully responsible on a political level.* Whereas the demand for a codified bill of rights may not be unreasonable the latter demand is fanciful and fails to appreciate the political nature and political potential of the Community. It is doubtful whether the Parliament will ever get full legislative powers. It can hope perhaps for co-decision powers. The Council of Ministers will not – unless fundamental changes in the Treaty take place – be subject to political control by the European Parliament.

[57] The French Cour de Cassation (Chambre Mixte) accepted the doctrine in the celebrated case *Administration des Douanes* v *La Société "Cafés Jacques Vabre" SA* (1975) 2 CMLR 336. The Court relied, however, on Article 55 of the French Constitution which gives Treaty provisions (subject to certain conditions, especially reciprocity) a force higher than French statutes, even those subsequently enacted. This reasoning does not amount then to full acceptance of the shift in the *Grundnorm.* It should be noted that Procureur Général Touffait had explicitly requested, in relation to Article 55 of the French Constitution, that the Court should not "... mention it and instead base [its] reasoning on the very nature of the legal order instituted by the Rome Treaty". This the Court implicitly declined to do. In a subsequent case, however, *Clave Bouhaten von Kempis* v *Geldolf* (husband and wife) [1976] Z *CMLR* 152, the Third Civil Chamber arguably "... did take the plunge..." (March-Hunnings, "Rival Constitutional Courts: A comment on Case 106/77", 15 *CML Rev.* 483 at 484 (1978)) and accepted the doctrine without reference to Article 55 of the French Constitution. The Criminal Chamber of the Cour de Cassation has also been Community minded – see, *Administration des Constitutions Indirects* v *Ramel* [1971] *CMLR* 357; and *Republic* v *Von Saldern et al.* noted in 10 *CML Rev.* 223 (1971). By contrast the Conseil d'État has, basing itself on somewhat antiquated notions of separation of powers, refused the acceptance of the supremacy principle as applied to parliamentary *loi.* The doctrine would probably apply to governmental decrees (cf. *Cohn-Bendit case,* note 8 *supra*). See *Syndicat Général des Fabricants de Semoules* [1970] *CMLR* 395. See also, *Syndicat des importeurs de vêtements et produits artisanaux CE* 28.5.1979, and CE 22.10.1979 [1980] *AJDA* 95. On the ambiguous position of the French Constitutional Court, see Mitchell, "What happened to the Constitution on 1 January 1973?", 2 *Cambrian L. Rev.* 69 (1980) at pp. 78-79 and notes therein. See also Kovar and Simon, "Some reflections on the decision of the French Constitutional Council of 30 December 1976", 14 *CML Rev.* 525.

[58] Ireland actually introduced a constitutional amendment – see Article 29, Third amendment to Irish Constitution of 1972. And see Temple-Lang, "Legal and constitutional implications for Ireland of adhesion to the EEC Treaty", 9 *CML Rev.* 167 (1972). The Danish Constitution already had a provision of delegation of powers to international organizations (Article 20 of Danish Constitution). But see, Due and Gulman, "Constitutional implications of the Danish accession to the European Communities", 9 *CML Rev.* 456 (1972), which analyses the debate as to the possibility of Danish compliance with the principle of supremacy especially *vis-à-vis* constitutional provisions (at pp. 265-267). There have been very few references from Denmark to the European Court so that the matter is still judicially open. See Rasmussen, "Survey of cases", 4 *EL Rev.* 484 (1979).

[59] See, for example, Wintertorn, "The British Grundnorm: Parliamentary supremacy re-examined", 92 *LQR* 591 (1976). But see Mitchell, note 57 *supra.*

[60] In the most recent case – indeed the first case in which the House itself made a reference under Article 177 – *R.* v *Henn & Darby* [1980] 2 *WLR* 597. Their Lordships did not raise directly the question of supremacy. They did, however, accept the duty to refer and by implication the binding authority of Community law as intepreted by the European Court. Note, however, that in any event this case was not concerned with British legislation subsequent to Community law. See, Faull, "Moralité publique et libre circulation des

produits", 4 *CDE* 446 (1980); Weiler, "Europornography, First Reference of the House of Lords to the ECJ", 44 *MLR* 91 (1981).

[61] Lord Denning's judicial statements have so oscillated that they must now be taken with a measure of caution. In *Blackburn* v *AG* [1971] 2 *All ER* 1380 he said "We have all been brought up to believe that in legal theory, one Parliament cannot bind another and that no Act (such as the European Communities Act, Sections 2 and 3 of which sought to entrench the supremacy of Community law) is irreversible... if Parliament should [try and revoke the Act], then I say we will consider that event when it happens", at p. 1382. It would seem thus that he was acknowledging that the sovereignty principle was a legal rule which could be changed by the Courts and that in the case of the Treaty of Rome that possibility of *Grundnorm* shift was not excluded. By contrast in *Felixstowe Dock and Railway Co.* v *British Transport Docks Board* [1976] 2 *CMLR* 655 he stated "It seems to me that once a Bill is passed by Parliament and becomes a Statute, that will dispose of all this discussion about the Treaty. These courts will then have to abide by the Statute without regard to the Treaty at all", at p. 664. In two subsequent cases his statements became more subtle. Thus in *Shields* v *E. Coomes (Holdings) Ltd* [1979] 1 *All ER* 456, he made the following hypothesis:
"Suppose that the Parliament of the United Kingdom were to pass a statute inconsistent with Article 119 [dealing with equal pay for women] by giving the right to equal pay only to unmarried women. I should have thought that a married woman could bring an action in the High Court to enforce the right to equal pay given to her by Article 119... If [the courts] should find any ambiguity in the statutes or any inconsistency with Community law, then [it] should resolve it by giving the primacy to Community law",
at p. 460. This may look like acceptance of supremacy but Denning was careful to choose a situation of inconsistency rather than conflict that is where the Community provision extended British law and did not directly conflict with it. But in *Macarthys Ltd* v *Smith* [1979] 3 *All ER* 325, he stated that filling the gaps in this way was on the assumption "... that our Parliament, whenever it passes legislation, intends to fulfil its obligations under the Treaty. If the time should come when our Parliament deliberately passes an Act with the intention of repudiating the Treaty or any provision in it or intentionally of acting inconsistently with it and says so in express terms then I should have thought that it would be the duty of our courts to follow the statute of our Parliament. I do not however envisage any such situation. As I said in *Blackbur* v *Attorney-General*: 'But if Parliament should do so, then I say we will consider that event when it happens'. Unless there is such an intentional and express repudiation of the Treaty, it is our duty to give priority to the Treaty",
at p. 329. With which Lawton LJ agreed.
The third judge (who subsequently retracted somewhat: cf. [1980] 3 *WLR* at p. 949) said "If the terms of the Treaty are adjudged Luxembourg to be inconsistent with the provisions of the Equal Pay Act 1970, European law will prevail over that municipal legislation", i.e. adopting a supremacy position although not in relation to subsequent legislation. This position does not change in the light of the final decision of the court of appeal once the reference from Luxembourg was received. Denning said then [1980] 3 *WLR* 947 at 949:
"... It is important now to declare – and it must be made plain – that the provisions of Article 119 of the EEC Treaty take priority over anything in our English statute on equal pay which is inconsistent with Article 119. That priority is given by our own law. It is given by the European Communities Act 1972 itself. Community law is now part of our law: and, whenever there is any inconsistency, Community law has priority. It is not supplanting English law. It is part of our law which overrides any other part which is inconsistent with it"

and:
"... Community law is part of our law by our own statute, the European Communities Act 1972. In applying it, we should regard it in the same way as if we found an inconsistency between two English Acts of Parliament: and the court had to decide which had to be given priority".
Denning could be read to be saying that the 1972 European Communities Act is no more nor less than any other Act and that supremacy in this case is simply a result of it being enacted after the 1970 Equal Pay Act which was at issue (although several provisions were re-enacted in the 1975 Act).
A leading authority summarizes the situation thus:
(1) While the European Communities Act 1972 remains in force, existing directly applicable or effective Community law will be the law in the United Kingdom, notwithstanding any legislation prior to the Act which is inconsistent with such directly applicable or effective law.
(2) Community law which is not directly applicable or effective will have no force in the United Kingdom until given effect by Act of Parliament or by order or regulation enacted pursuant to powers given by the European Communities Act 1972 or other legislation.
(3) Directly applicable or effective Community law will take effect in the United Kingdom, notwithstanding any legislation prior to the 1972 [European Communities] Act, or even after the Act but prior to the coming into effect of the Community rule.
(4) A subsequent Act of Parliament which is inconsistent with a rule of Community law will be read subject to the rule of construction in s. 2(4) so that Community law can take effect notwithstanding the Act, at any rate if the Court is satisfied that the subsequent inconsistent legislation is not intended, expressly or impliedly, to repeal s. 2(1) or (4) of the 1972 [European Communities] Act; but
(5) Any subsequent Act of Parliament inconsistent with the European Communities Act 1972, including one which repeals the latter in whole or in part and one which is intended to limit the application of s. 2(1) and (4), will be given effect by the United Kingdom courts.
L. Collins, *European Community law in the UK*, Butterworhs, London, 1980, at pp. 25-26. But see note 153 *infra*.

[62] Note 41 *supra* at p. 7 (emphasis supplied).

[63] Idem, at pp. 23-24.

[64] I am most indebted here to a paper delivered by Professor M. Waelbroeck at a conference held in Bellagio, Italy in July 1979 entitled "Community pre-emption and related problems" on which I have relied extensively in this part. See also D. J. Gijlstra et al. (eds.), *Leading cases and materials on the law of the European Communities*, Kluwer, Deventer, 1977, at pp. 88-95.

[65] Thereby indirectly prohibiting national legislation even if compatible in fields already occupied by a regulation. See for example, Case 34/73 *Variola* [1973] ECR 990; Case 39/72 *Commission* v *Italy* [1973] ECR 113.

[66] Case 22/70 *Commission* v *Council* [1971] ECR 273.

[67] Recitals 17-18 of Judgment, p. 274.

[68] Opinion 1/75 [1975] ECR 1335.

[69] Idem at p. 1363. Note, however, that unlike ERTA where there were no Treaty provisions at all for external competence in relation to transport, the OECD case fell within the commercial policy which has such provisions.

[70] Joined Cases 3, 4 and 6/76, *Cornelis Kramer and Others* [1976] ECR 1279.

[71] Idem, Recitals 35-39 of Judgment, p. 1310.

[72] Treaty of Accession, Article 102.

[73] Opinion 1/78 [1979] ECR 2871.

[74] Recitals 57-60 of Judgment, pp. 2917-2918.

[75] See note Gijlstra, note 64 *supra*.

[76] Hanson, "Methods of interpretation - A critical assessment of results", in Judicial and Academic Conference, Court of Justice of the European Communities, Luxembourg, 1976, II.

[77] Idem at II 9; II 25 and II 26. Hanson's critique is not merely one of judicial policy and judicial role. In analysing *Van Gend en Loos* he makes an acute distinction between *categorical* Treaty provisions which prescribe a particular consequence (e.g. Article 85 (2)) and *imperative* provisions which only prescribe a legal obligation. He maintains that the doctrine of direct effect as expounded in *Van Gend en Loos* and its progeny renders – unjustifiably and illegitimately – all imperative provisions categorical.

[78] Idem at II 25.

[79] Whereby, unlike the classical condominium in which there is joint national control over territory, here there is joint control over economic and other sectors of public policy.

[80] Cf. Tenth Amendment to the US Constitution.

[81] On the doctrines of implied and additional powers, see, for example Gijlstra, note 64 *supra* at pp. 1-9. See also, Schwartz, "Article 235 and law-making povers in the European Community", 27 *ICLO* 614 (1978) and Mitchell, note 55 *supra* at p. 73. The discussion on implied or additional powers is crucial in fully-fledged federations like Canada and the USA since two sets of *independent* power centres exist. The issue becomes much less acute in the Community since it is the governments of the Member States which control the legislative process of the Community. Expanding the Community jurisdiction will thus *ipso facto* be by the consent of at least a majority of Member States. Whether expansion can be done through elastic clauses such as Article 235 or necessitates actual Treaty amendment is less important except in as much as *individuals* may contest the expansion of Community jurisdiction undertaken by their government or in ERTA situations where the Member States prefer to proceed outside the Community framework.

[82] See, for example, Case 22/70 note 64 *supra* (external relations - transport policy); Opinion 1/78, note 71 *supra* (common commercial policy); Case 91/79 *Commission* v *Italian Republic* [1980] ECR 1099 (environmental policy). See also Gijlstra; Schwartz note 81 *supra*.

[83] There is scope to examine the reasons for the totally different approach of, say, the Privy Council – for many years the "supreme court" of Canada – in relation to the Canadian Constitution. Was the restrictive Privy Council approach a result of a reasoned assessment of the different character of the Canadian federation, or perhaps the result of the British Law Lords' inability to adapt to the different spirit of federal jurisprudence?

[84] Cf. Hay, note 15 *supra*, at p. 69 ff. Mitchell, note 57 *supra*, at p. 73. Objections can of course be raised by national parliaments (see, for exaple, Debate in British House of Lords, 4 July 1978; see also note 8 *supra*) and national pressure groups, although the former have, at least in theory, the ability to control their ministers' activity at the Community level. The "political estoppel" imposed on the governments of the Member States in challenging jurisdictional expansion was recently illustrated in Case 91/79, note 82 *supra*. In that case Italy was defending its non-implementation of Directive 73/404/EEC which formed part of the Community's environmental programme. Although serious doubts exist as regards the Community's competence to operate in this field – (cf. E. Grabitz and C. Sasse, *Competence of the European Community for environmental policy* (1977), at pp. 24-31) – and therefore about the legality of the Directive in question, the Italian Government which partook in evolving the policy (see, OJ C 12, 20.12.1973) could not but state in that case that it did "... not intend to raise the question whether the directive is valid in the light of the fact that combating pollution is not one of the tasks entrusted to the Community by the Treaty", at p. 1103.

[85] This is not to imply the Court's role is a political. In fact the Court has shown a high degree of political acumen in, say, changing course on the question of human rights. It has also played an important role in demarcating the division of competences between different European Community institutions. See, for example, Case 22/70 *Commission* v *Council* [1971] ECR 263 and Joined Cases 138/79; 139/79, note 5 *supra*.

[86] This interplay is well illustrated in C. Sasse, E. Poullet, D. Coombes, G. Deprez, *Decision-making in the European Communities*, Prager Publishers Inc., N.Y. - London, 1977, and H. Wallace, W. Wallace, C. Webb (eds.), note 34 *supra*.

[87] See Wallace et al., note 34 *supra*, especially Webb, "Introduction: Variations on a theoretical theme", at pp. 1-31; and H. Wallace, "National bulls in the Community china shop: The role of national governments in Community policy-making", at. pp. 33-68.

[88] This inevitable feature of Community life is not only a sharp reminder that it is still a *Europe of nations* but also places quality restraints on the body. "The need to cater for the differing interests of the Member States and to accept national quotas, however unofficial, ...inhibits the development of an elite corps of policy-makers...", H. Wallace, idem, at p. 53.

[89] Article 10, Merger Treaty.

[90] It should be noted that apart from a few limited fields such as competition, the actual practical execution of many Community policies is in the hands of the Member States acting as agents for the Community.

[91] See Pescatore, note 32 *supra*, at pp. 7-10.

[92] See Article 149 EEC.

[93] There is here an apparent paradox. To the extent that, say, the governments of all 10 Member States agree on a "desired" course of action, how can this be said to be contrary to the Community spirit? For do not the governments represent the Member States which together *are* the Community? True as this may be, the originating Treaties still remain the main normative basis for Community evolution. To the extent that the governments consider disregarding the Treaty objectives (without formally amending it) they may legitimately be characterized as acting contrary to the Community spirit.

[94] There is a certain terminological confusion regarding "secondary" legislation. Regulations and directives are often referred to as secondary legislation. The better view, it is submitted, is to regard them as primary legislation. For if we view the Treaty as being the Constitution of the European Communities, the legislation thereunder, by analogy to national systems, would be primary. Perhaps the power of legislation entrusted to the Commission under enabling measures of the Council may be characterized as "secondary".

[95] A recent study giving a realistic appreciation of the institutional balance is S. Henig, *Power and Decision in Europe*, Europotentials Press, London, 1980; see also Sasse, note 86 *supra*.

[96] Term borrowed from Marquand, note 4 *supra*.

[97] The following remarks by a leading commentator are instructive in this context. "...All Member States have organized their policy-making in such a way as to promote their own national interests... These efforts to keep the formation and implementation of Community rules under national control are sustained by the fact that the organs of the European Communities still lack a democratic legitimation of their own... To date, the control of European policy through national parliaments is at any rate comparably weak and is at most exerted via a detour that is through the control of governments". Sasse, "The control of the national parliaments of the Nine over European affairs", in A. Cassese (ed.), *Parliamentary control over foreign policy*, Sijthoff & Noordhoff, Alphen aan den Rijn, Germantown, 1980, at p. 147.

[98] In fact a different type of democracy deficit emerged whereby the executive branch of the governments of the Member States able to pursue policies which, perhaps, would be more

open to challenge if conducted "at home". See Rogers & Bolton, "The true Cost of Sanctions", *The New Statesman*, 9.1.1981.

[99] Cf. Case 1/58 *Friedrich Stork & Co.* v *High Authority of the European Coal and Steel Community* [1959] ECR 17.

[100] See Cappelletti, "The mighty problem of judicial review and the contribution of comparative analysis", *LIEI* 1 (1979). On recent developments of human rights and the Community legal order see, Economides and Weiler, "Accession of the Communities to the European Convention on Human Rights: Commission Memorandum 42", *MLR* 683 (1979). See also, Case 44/79 *Liselotte Hauer* v *Land Rheindland-Pfalz* [1979] ECR 3727.

[101] See H. Wallace, note 87 *supra*, at p. 53.

[102] See, for example, "Report on European institutions presented by the Committee of Three to the European Council" (Council of the European Communities, Luxembourg, 1980), hereinafter *Three Wise Men Report*, pp. 49-53.

[103] On this effect of Coreper the Three Wise Men commented that the Commission "... should not, as so often happens now, be drawn into negotiating with [national experts, etc.] to find a supposedly acceptable form of the [policy] measure". They also commented that "[the] Commission must frame its proposals in a more *independent* manner". *Three Wise Men Report*, p. 54.

[104] *Three Wise Men Report* at, for example, 53 and *passim*.

[105] For texts and brief commentary conférence de presse du Président de Gaulle, 9 September, 1965; Dichiarazione del Consiglio dei Ministri delle CEE, Bruxelles, 26.10.1965; Communiqué de presse sur les accords de Luxembourg, 29.1.1966. All in R. Ducci, B. Olivi, *l'Europa Incompiuta* 411-422 (Padova, Cedam, 1970). See also P. J. G. Kapteyn, P. Verloren van Themaat, *Introduction to the Law of the European Communities*, London, Sweet & Maxwell, 1973, at pp. 143-146.

[106] For example, the procedure in the Agricultural Management Committees. See text to note 114 *infra*.

[107] For example, Article 126 EEC - "[The] Council, after receiving the opinion of the Commission..., may... unanimously determine what new tasks may be entrusted to the [European Social Fund]".

[108] *I.e.* where the Treaty provides for a qualified majority in Council a small number of States can veto a decision.

[109] See, OJ 71, 4.11.1961; OJ L 326, 29.12.1969. In relation to agreements based on Article 238 EEC which does not provide for a consultative committee, the Council set up a group of national observers who "constitute... more than token representation, although less than active intervention". But, "Negotiating Community agreements: Procedure and practice, *CML Rev.* 286, at p. 294 (1970).

[110] Among commentators there is no dispute about the decline in the Commission role although opinions differ about the appraisal of the phenomenon. Costanis, "The treaty-making power of the European Economic Community: The perspective of a decade, 5 *CML Rev.* 421 (1967-8), Alting von Geseau, "The external representation of plural interest", 5 *Journal of Common Market Studies* 426 (1967), and Leopold, "*External relations power of the EEC in theory and practice*", *ICLO* 54, especially at pp. 59-62 (1977) tend to be more critical of this development. Bot (idem) offers a more balanced approach maintaining that the exigencies of international relations and their conduct make this development largely inevitable.

[111] Costanis, idem, p. 435.

[112] Idem, p. 434 (emphasis added). But see Bot, note 109 *supra*, at pp. 309-310. For an amusing account of the technical obstacles which this negotiating pattern produces, see W. Field, *The European Community in world affairs*, Alfred Publishing, Washington, 1976, at p. 105.

[113] See generally Weiler, "*The European Parliament and foreign affairs: External relations of the EEC*", in Cassese, note 97 *supra*, at p. 151.

[114] See *Council and Commission Committees*, Supplement 2/80, Bull. EC, Commission of the European Communities, Luxembourg, 1980, at pp. 18-19.

[115] Idem at p. 22.

[116] The Court of Justice affirmed the legality of management committees and even seemed to approve their political role. In a case before the Court it was argued first that since execution of policies involved legislative functions it should have been left entirely in the hands of the Council. The Court first acknowledged the distinction: "according [with] the legal concepts recognized in all Member States, between... measures directly based on the Treaty itself and derived law intended to ensure their implementation. It cannot [the Court concluded] therefore be a requirement that all the details of the regulation concerning the common agricultural policy be drawn up by the Council..." An alternative argument was that the "... management committee procedure... constituted an interference in the Commission's right of decision, to such an extent as to put in issue the independence of that institution. Further the interposition between the Council and the Commission of a body which is not provided for by the Treaty is alleged to have the effect of distorting the relationships between the institutions and the exercise of the right of decision". The Court's reply was that the management committee procedure was a legitimate exercise by the Council of its power under Article 155 to stipulate conditions under which it would delegate power. Thus *per* the Court: "Without distorting the Community structure and the institutional balance, the management committee machinery enables the Council to delegate to the Commission an implementing power of appreciable scope, subject to its power to take the decision itself if necessary". Case 25/70 *Einfuhr- und Vorratsstelle* v *Köster* [1970] II ECR 1161 at 1170-1171; Recitals 3-10 of Judgment.

[117] *Accord* - Doc 115, 1968-69 at 30.9.1968 of European Parliament and Resolution of 3.10.1968, OJ C 108, p. 37.

[118] See generally *Proposals for reform of the Commission of the European Communities and its services, Report made at the request of the Commission by an Independent Review Body Under the Chairmanship of Mr Dirk Spierenburg*, Commission of the European Communities, Brussels, 1979.

[119] A recent event has been the intervention in January 1981 by the British Prime Minister with the President of the Commission as regards the portfolio of one of the British Commissioners.

[120] The so-called "mixed agreement". See Kapteyn & Verloren Van Themaat, note 105 *supra*, at p. 351 ff.

[121] Shonfield, note 11 *supra*.

[122] Idem at p. 10.

[123] Case 48/71 *Commission* v *Italian Republic* [1972] ECR 532; Case 48/69 *Commission* v *ICI Ltd* [1972] ECR 649.

[124] Shonfield, note 11 *supra*, at p. 17.

[125] See notes 7-8 *supra*.

[126] Akehurst, "Withdrawal from International Organizations", 32 *Current Legal Problems* 143 (1979).

[127] The leading authority on withdrawal generally is Feinberg, 39 *BYBIL* 189 (1963) on whom Akehurst draws.

[128] Akehurst, idem at p. 151. A view shared even by "... international lawyers from Communist countries, which normally argue that there is an implied right of withdrawal from international organizations", idem. This accords of course with the European Court of Justice view. See Case 128/78 *Commission* v *UK* [1979] ECR 419, especially at p. 429.

[129] Idem.

[130] Pryce, note 21 *infra*, at p. 55. Since the *travaux* of the Treaties of Paris and Rome have not been released it is difficult to see how Pryce can be so assertive in his submission.

[131] Idem.

[132] "Alliance politics" is the refinement of the "bag of sticky marbles" concept. See A. Shonfield, *European integration in the second phase: the scope and limitation of alliance politics*, The University of Essex - Noel Buxton Lecture, 1974.

[133] See Feinberg, and Akehurst, notes 127, 126 *supra* and examples therein. Indeed, in the few cases cited "withdrawal" was subsequently construed as "suspended membership".

[134] And see generally, A. W. Green, *Political Integration by Jurisprudence*, Sijthoff, Leyden, 1969.

[135] See "*Case Concerning United States Diplomatic and Consular Staff in Tehran (United States of America* v *Iran)*" [1980] *ICJ* (19 ILM 555 (1980)).

[136] See M. Cappelletti, *Judicial review in the contemporary world*, Bobbs-Merrill, Indianapolis, 1971, especially, Chapter 4; and Cappelletti, "*Giustizia costituzionale soprannazionale*", 23 *Rivista di diritto processuale* 1 (1978).

[137] On the system of juridical review generally, see e.g. the erudite treatment by Schermers, note 55 *supra*; L. Neville Brown & F. G. Jacobs, *The Court of Justice of the European Communities*, Sweet & Maxwell, London, 1977. G. Vandersanden & A. Barav, *Contentieux communautaire*, Bruylant, Brussels, 1977. Naturally my treatment here will only sketch the bare limbs of the complex system.

[138] See Schermers, *idem* § 53 - § 144 (pp. 31-80).

[139] In this context one may record the words of Justice O. W. Holmes: "I do not think the US would come to an end if we lost our power to declare an Act of Congress void. I do think the Union would be impended if we could not make that declaration as to the laws of the several States". *Collected Legal Papers*, pp. 295-296, NY, Peter Smith, Reprint 1952.

[140] E.g. Articles 173, 175, 184, 215 EEC.

[141] Cf. Stein & Vining, "*Citizen access to judicial review of administrative action in a transnational and federal context*", in F. Jacobs, *European law and the individual*, North-Holland, Amsterdam, 1976, p. 113.

[142] *E.g.* Articles 169, 170 EEC.

[143] Thus the Commission has apparently decided not to bring an action against France in the wake of the decision of the Conseil d'État in the *Cohn-Bendit case*, note 8 *supra*, although a clear violation had taken place (see, Isaac, "*L'affaire Cohn-Bendit*", 15 *CDE* 265, 1979). The Commission may fear that bringing an action under Article 169 EEC might give the impression of interfering with the independence of the French judiciary. This attitude, if it is the basis of the Commission's reluctance, is, it is submitted, unconvincing. The major part of the jurisprudence on the *European Convention on Human Rights* is dependent on the exhaustion of local remedies which implies *ipso facto*, the subjection of judicial decisions to scrutiny by the organs of that Convention. The same would apply *mutatis mutandis* to "169 actions" based on wrong decisions taken by Member States' final appeal courts.

[144] See Commission observations in Case 144/77 *Commission* v *Italian Republic* [1978] ECR 1307 and see text to note 163 *infra*.

[145] Idem.

[146] See note 8 *supra*.

[147] Jacobs, "*When to refer to the European Court*", 90 *LQR* 486 (1974), at p. 486.

[148] See text to note 212 *infra*.

[149] The substantive review grounds for *legality* under Art. 173 EEC and *validity* under 177

are largely similar although certain slight differences exist. See Schermers note 55 *supra* at § 371 (pp. 210-211).
The breadth of the remedy is also different. Under 177 which is, strictly speaking, an inter-court procedure the individual's right of arguing his case is more limited. In addition since the time limit for annulment under 173 will have passed it would be strange if *ex tunc* effect could be given to the 177 reference unless specifically indicated by the Court of Justice. *Cf.* Bebr, *Remedies for Breach of Community Law*, Community Report to 9th Congress of FIDE (FIDE & Weet & Maxwell, London, 1980) at 10.5-10.6. *See* note 151 *infra*.

[150] Case 61/79 *Amministrazione delle Finanze dello Stato* v. *Denkavit Italiana*, Judgment of Court of 27.3.1980 (not yet reported) Recital 12 of Judgment.

[151] Rasmussen "*Why is Article 173 Interpreted Against Private Plaintiffs?*" 5 *EL Rev.* 112 at 115 (1980).
Rasmussen's thesis, if correct, illustrates another subtle evolutionary trend in the very system of judicial review. It is not proposed to discuss here all the various advantages and disadvantages that each limb of the system entails. It has e.g. been justifiably pointed out that the 177 procedure for judicial review of Member State actions does not allow the Member States sufficient facilities (by comparison to 169-170 EEC procedures) to defend their position before the European Court. *See, UK House of Lords Select Committee on the European Communities (1979-80) House of Lords 23 Report*, HMSO). See generally, Rasmussen, *idem* and comprehensive bibliography cited in note 6 therein.

[152] *Wisconsin* v. *Constantineau*, 400 US 433, 436 (1971).

[153] There are two elements in this cooperation between the European Court and national courts. First, the national courts must show a willingness to make references. The pattern is checkered. In 1979 the following references were made: *Belgium*: 6 from the Cour de Cassation, 7 from courts of first instance or of appeal; *Denmark*: 1 from a court of first instance; *France*: 2 from the Cour de Cassation, 2 from the Conseil d'État, 14 from courts of first instance or of appeal; *Federal Republic of Germany*: 2 from the Bundesgerichtshof, 1 from the Bundesverwaltungsgericht, 9 from the Bundesfinanzhof, 5 from the Bundessozialgericht, 16 from courts of first instance or of appeal; *Ireland*: 1 from the High Court, 1 from the Chuirt Chuarda; *Italy*: 7 from the Corte di Cassazione, 12 from courts of first instance or of appeal; *Luxembourg*: 1 from a court of first instance; *Netherlands*: 1 from the Hoge Raad, 1 from the Centrale Raad van Beroep, 3 from the College van Beroep voor het Bedrijfsleven, 1 from the Tariefcommissie, 5 from courts of first instance or of appeal; *United Kingdom*: 1 from the House of Lords, 1 from the Court of Appeal, 6 from lower courts. For a useful attempt at analysing the different pattern of references, *see* Mortelmans, "*Observations in the Cases Governed by Article 177 of the EEC Treaty: Procedure and Practice*", 16 *CML Rev.* 557 (1979).
Secondly, once the case is remitted back to the national court the latter must construe the case – where discretion is given – in a manner consistent with and truthful to the European Court's ruling. The English Court of Appeal (on receipt of the European Court's judgment in *Santillo* note 147 *infra*, *R.* v *The Home Secretary, ex parte Santillo*, The Times, 23 December 1980) "presided over with the pertinacious idiosyncracy" of Lord Denning MR illustrated how, if it "... approves of a rule of English law, no amount of EEC law will deflect [it] from giving the former priority". "*European legal movement shows signs of strain*", Financial Times, 22.12.1980 at p. 16.

# III

# Economics

# The Optimum Monetary Constitution: Monetary Integration and Monetary Stability

EMIL-MARIA CLAASSEN
Paris/Florence

## I. Introduction

Over the last ten years, there has been (again) in the literature an increasing interest in monetary reform, not in terms of a better conduct of monetary policy (a subject of constant concern over the last decades), but in terms of a better institutional framework of the monetary system. We shall call this institutional framework the monetary constitution of an economy. We have borrowed the notion from the relatively unknown, but fundamental monograph by Friedrich Lutz on *Das Grundproblem der Geldverfassung*, published exactly fifty years ago.[1]

One possible reason for the increased scholarly interest in monetary reform is that of the inflationary experience in the industrialized countries over the 1970s. This inflationary decade coincides with the emergence of a world monetary system which, according to Milton Friedman (1985, p. 2), is "unprecedented". The year 1971, the end of Bretton Woods, is the formal date for the removal of the remaining relics of a bygone commodity standard. Even though, sind World War I, the world monetary system moved already gradually toward a dollar standard, some elements of a commodity standard had still been maintained: fixed exchange rates with respect to the dominant currency – the dollar, and the pretense of the U.S.A. that the dominant currency is based on a gold standard. Thus, since the 1970s, the world monetary system has given up the last formal links to a centuries-old commodity standard and it is now definitely on a fiat standard, i.e., in Irving Fisher's terms, on an irredeemable paper money standard.

The new era of a worldwide fiat standard was accompanied, over the 1970s, by a more than doubled world inflation rate (where the term world stands here for the Western industrialized countries). Whether this inflationary phenomenon was due to an argument of the type *post hoc, ergo propter hoc* or not, our profession was looking for alternative monetary systems or for the improvement of the actual monetary system which would prevent more efficiently inflation. The better conduct of (a discre-

tionary or automatic) monetary policy or the return to fixed exchange rates or/and to some commodity standard were one of the main preoccupations. Another and a completely new strand of ideas was initiated mainly by Friedrich Hayek's proposal for a Free *Choice in Currency* (1976a) and for the *Denationalisation of Money* (1976b). While in the "older" literature, the monopoly of currency (base money) creation by government was considered as a necessary condition for price stability, this argument was reversed by Hayek suggesting an unrestricted supply of private high-powered monies:

> The past instability of the market economy is the consequence of the exclusion of the most important regulator of the market mechanism, money, from itself being regulated by the market mechanism." (Hayek, 1976b, p. 79.)

In the recent literature on the competitive money supply, the performance of a competitive versus a monopolistic system of money creation is judged on the grounds of monetary stability. However, our concern is to enlarge the efficiency criteria of a monetary system. Money is a store of value *and* a means of payments. The highest degree of moneyness is achieved when money is a perfect store of value (monetary stability) *and* a generally accepted means of payments (monetary integration). The goals of a monetary constitution are the realization of monetary stability and monetary integration within the framework of a country or with respect to several countries or even in regard to the whole world economy (section II). The institutional means and, in particular, the question of whether government intervention is a necessary condition for achieving the two fundamental objectives of a monetary constitution and, thus, for bringing about the "optimum" monetary constitution are considered, successively, for the realization of monetary integration (section III) and of monetary stability (section IV).

The methodology of our arguments will be rather abstract. In section III we ask the question whether there can be made a case for an invisible hand emergence of a money economy from a barter economy and of full monetary integration from a low level of monetization. Section III concerns the demand side of money, i.e. its use and the demand for money holdings, while the supply side of money is discussed in section IV. One of the main purposes of section IV is to know whether a monopolistic or competitive system of money creation is the best regime for achieving monetary stability and, in particular, whether a competitive fiat standard is viable in comparison with a commodity standard.

## II. The Fundamental Objectives of the Monetary Constitution

Each monetary constitution has (or should have) two objectives: monetary integration and monetary stability. Monetary integration aims at improv-

ing the quality of money as a generalized *means of payments*, whereas monetary stability stresses the other function of money which is the *store-of-value* function.

First of all, we must specify what one could understand by monetary integration and monetary stability. As we shall see, there are different degrees of monetary integration and different interpretations of monetary stability (Claassen, 1984, pp. 49-50).

*Different degrees of monetary integration.* The highest degree implies the existence of a single currency in the world economy. The other extreme representing the lowest degree of monetary integration is more difficult to define. It could be a world of n countries with n currencies related to each other by flexible exchange rates. In such a case, a world of n currencies with fixed exchange rates would be very near to the hypothetical one-currency world, so that a regime of fixed exchange rates and a regime of flexible exchange rates reflect the different degrees of monetary integration: the lower the degree of flexibility of the exchange-rate system, the higher the degree of monetary integration within the world economy.

However, one can also imagine an even lower degree of monetary integration by assuming that there are more than n currencies within a system of flexible exchange rates. Thus, for instance, each region of a country may have several currencies which may circulate with the currencies of other regions and even with those of other countries. In this situation the lowest degree of monetary integration would have been achieved.[2]

The number of currencies and the degree of flexibility of their exchange rates have been chosen as the criteria for the degree of monetary integration. If one takes the number of producers as the classification scheme for the degree of competition in the economy (by assuming that the demand side is atomistic), the highest degree of monetary integration would imply a monopolistic monetary system and a low degree of monetary integration would involve an oligopolistic or competitive monetary system.

*Different interpretations of monetary stability.* At first sight, monetary stability should not evoke any definition problems because it could be specified by a constant general price level (even though the latter may imply statistical difficulties of a well-defined price index). Consequently, a variable price level, upwards and downwards, would represent the case of monetary instability, and special cases of this monetary instability are inflation and deflation.

We know from the literature that inflation is harmless for the economy provided that it is correctly anticipated: Under this condition, the real variables of the inflationary economy are identical to those of an economy with a constant price level except the level of real cash balances which will be reduced; however, this reduction of real cash balances could be avoided by the payment of an interest rate on money holdings by the amount of the

inflation rate. Consequently, the target of monetary stability in terms of a constant price level would not represent fundamentally higher advantages for the economy than the inflationary regime in which money yields an interest rate of the amount of the inflation rate. The case for the absolute superiority of a constant price level can only be made on the grounds of the existence of wrong inflationary expectations. However, even monetary stability in terms of a constant price level does not necessarily involve correct expectations with respect to the future price level because we never know exactly what will be the price level in the future.

When, for reasons of *better* (i.e. more correct) expectations, a constant price level can be chosen as indicator of monetary stability, it is not entirely evident that deflation is not the best solution for realizing monetary stability because deflation may *also* yield the optimum quantity of money. But we shall (or we must) distinguish between the objective "monetary stability" and the objective "optimum quantity of money".

One could argue that the monetary constitution should realize *three* objectives: the highest monetary integration (a world currency), the highest monetary stability (a constant general price level) and the optimum quantity of money (the payment of an interest rate on money equal to the real market interest rate). However, our attention is concentrated more on the two first objectives since the optimum quantity of money could always be realized whatever may be the actual degree of monetary integration and of monetary stability.

*Trade-off between monetary integration and monetary stability.* Both objectives can be incompatible with each other over certain periods. Thus, for instance, higher monetary stability may be attained by the introduction of a flexible exchange rate which implies a lower degree of monetary integration; vice versa, a higher degree of monetary integration through a fixed-exchange-rate system or via the formation of a currency area may involve a lower level of monetary stability. It should be emphasized that this possibility of a trade-off relationship has to be seen in a long-run perspective (or in a historical retrospective). Consequently, the whole argument of this paper is based in terms of decades (and even centuries). Applied to the monetary scene of Europe it could explain the different stages of a higher or lower degree of the European monetary integration, in the past and probably also for the future.

## III. Monetary Integration: The Economies of the Use of Money and its Institutional Implementation

The *use* of money, in contrast to the production of a certain quantity of money, as a unit of account and a medium of exchange represents a revolutionary technique of organizing the exchange of goods in a market economy. This technical progress in the organization of exchange means a

considerable reduction in transaction costs associated with the buying and selling of goods. The economies of transaction costs due to the existence of money can be specified by assuming successively a higher monetization degree – or a higher degree of monetary integration – of an economy. The first "quantum jump" in the economies of transaction costs arises from the transition from a barter economy to a money economy. However, the transformation into a money economy needs not to imply a fully monetized economy, and the second decline in transaction costs results from the generalized use of money as a unit of account and a means of payments. Thirdly, several monies could still be used as a unit of account and a medium of exchange, and the "monetary unification" stands for the third type of economies of transaction costs.

The three different kinds of economies in transaction costs could have, each of them, an internal and external component. To the extent that there are non-pecuniary externalities, one could ask the question of whether the monetization of the economy is fully realizable by the private sector of the economy or whether government intervention is a necessary condition for the accomplishment of the various degrees of monetary integration.

## 1. The Economies of the Transition of a Barter Economy Towards a Money Economy

The size of the economies in transaction costs of a money economy depends on the type of a barter economy with which one compares the money economy. By assuming an already *organized barter economy*, the latter can be conceived in terms of trade posts or in terms of trade intermediaries.

*Trade posts.* One possible organization form for the exchange of n goods within a barter economy is the establishment of n(n-1)/2 trade posts. By assuming, for sake of simplicity, the existence of four commodities, A, B, C and D (where n = 4), there would be six trade posts. The implementation of this trade pattern is comparable with a traffic system between n villages (Wallace 1972, pp. 838-39) where each village is connected to all other villages by a particular road (Fig. 1). Thus, for village A, there are three different roads connecting A with B, C and D. The total number of roads are six. In analogy, commodity A can be exchanged against commodities B, C and D, respectively. Thus, the resulting six trade posts are AB, AC, AD, BC, BD, CD.

The superiority of a money economy with respect to a barter economy being organized in terms of trade posts, is not evident at first sight. On the one hand, in a money economy, one would have four specialized trade posts (or markets) for each of the four commodities compared with six

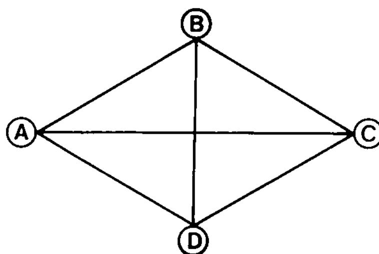



*Figure 1*

trade posts in a barter economy. Thus, for instance, instead of exchanging directly A against B at the trading post AB, in a money economy A has to be sold in the market of commodity A against money and B has to be bought in the market for commodity B against money.

From the point of view of a less costly organization of exchange, a money economy will only emerge from a critical number of commodities to be exchanged. In an economy with 10 commodities, a barter economy of 45 trade posts could still be associated with lower transaction costs than a corresponding money economy with 10 markets. However, when the number of commodities goes up to 100 goods, the lower transaction costs associated with only 100 specialized markets in a money economy compared with 4950 trade posts in a barter economy become more evident. Consequently, the economies of transaction costs emerge from a critical number of commodities onwards and then they rise, absolutely and relatively, with an increasing number of commodities.

It should be mentioned that the advantages of a money economy are often derived from the disadvantages of a non-organized barter economy. In the latter type of economy, there are two further kinds of transaction costs. On the one hand, there are search costs as a result from the search for the location of potential buyers or sellers. Even after a very intensive search process, the desired exchange between buyers and sellers does not need to coincide such that a complicated chain trade of "intermediary commodities" emerges. On the other hand, there are information costs about the quality of the traded goods (Alchian, 1977). Since there are no specialized agents, each buyer or seller must rely on his own knowledge about the quality of the supplied and demanded goods. Only specialized markets – in form of trade posts in an organized barter economy or of markets for one good in a money economy, respectively – permit an "objective" information about the quality of goods.

How could one conceive that a barter economy of the non-organized

type moves gradually toward a money economy directly, or indirectly by passing first via an organized barter economy? Are there sufficient market forces (the "invisible hand") which bring about a money economy or is an external enforcement or coercion (e.g. via government or some collective decisions) necessary to produce a more efficient, i.e. less costly, mechanism of exchange? If one looks through the glasses of traditional Walrasian general equilibrium theory, one could come to the conclusion that there should be some central decision maker who organizes the exchange of goods. Walras' fictitious auctioneer, with his concepts of "tâtonnement" and "prix créés au hasard" in a world without money, who realizes the equilibrium price at which the demand for each good is equal to its supply, has still not solved the exchange mechanism of goods (Starr, 1972; Barro and Fischer, 1978, pp. 17-19). His function could be augmented by implementing an organized barter economy (e.g. of the above kind of trade posts) or by introducing a general medium of exchange. However, as far as the organization of a barter economy is concerned, it should be considered as rather self-evident that buyers and sellers have a self-interest to introduce such a technical trading innovation since the transaction costs (search costs of location, information costs on quality) decrease sharply. Another question is whether a money economy, out of organized or even non-organized barter economy, will not evolve because of the fact that transaction costs (in terms of the number of "markets") would even be reduced more radically. The evolution of a money economy can be exemplified better by assuming an organized barter economy dominated mainly by trade intermediaries.

*Trade intermediaries.* This type of an organized barter economy is comparable with a traffic system which connects the different villages (B, C, D in Fig. 2a) with a centre village (A) serving as junction. In our example of a barter economy with four goods and assuming four economic subjects, the following desired pattern of exchange could be taken as an example:

economic subject I supplies A and demands D;
economic subject II supplies B and demands A;
economic subject III supplies C and demands B;
economic subject IV supplies D and demands C.

One possible trade configuration is that of Fig. 2b in which the economic subject I is simultaneously the trade intermediary. While all other economic subjects undertake a bilateral exchange with I, the latter performs a multilateral exchange with all others.

It should be noted that each of the four economic agents could exercise the role of an intermediary. The one who will overtake the function of the trade centre – in our example the economic subject I – is that agent who offers the best trade services in terms of the knowledge about the location

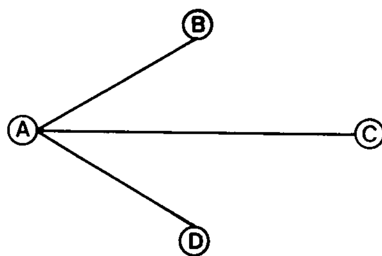



Figure 2a

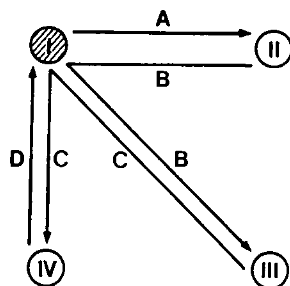



Figure 2b

of buyers and sellers, about the quality of the demanded and supplied goods and about possible price differences. The trade intermediary is, to a certain extent, the trade post for all goods. With an increasing number of goods, it is highly probable that there is an emergence of several trade intermediaries which become specialized geographically and/or in certain types of goods. And again, as in the case of trade posts, from a certain critical number of goods, it seems more "rational" (i.e. less costly) that a money economy emerges.

The emergence of a money economy may be related intimately with the existing trade centre(s). Thus, it is possible that the economic agents choose gradually as an "intermediate commodity of exchange" a commodity in which the trade intermediary is particularly specialized or for which the trade intermediary has created a perfect market in terms of the perfect certainty about the quality and the price of the concerned commodity (besides other traditional attributes like homogeneity, diversibility, transportability . . .). *After* the market participants having selected *progressively* the most convenient commodity as the unit of account and the medium of exchange, historically government was tempted to step in for certifying additionally the quality of the commodity having been chosen as money:

> Historically, a single unit of account linked to a single dominant . . . money has tended to emerge, initially via a market process of transactors settling on a particular commodity, followed almost invariably by government's exercising over one or more aspects of the issuance of . . . money – typically with the ostensible purpose of standardizing the coinage and certifying its quality (purity, fineness, etc.). Occasionally, two commodities, with a flexible rate of exchange between them, have simultaneously been . . . monies, one for small transactions, the other for large, as with silver and gold in the Middle Ages, or copper and silver in China. *(Friedman and Schwartz, 1985, p. 10.)*

As Carl Menger (1982) has emphasized, there is a clear-cut case for the invisible hand explanation of the genesis of money from the initial state of a barter economy since it is in the interest of each trader to diminish the

transaction costs associated with barter, even if the barter economy is already of the organized type (in form of trade posts *and* trade intermediaries). The "spontaneous" emergence of money concerns the unit of account as well as the medium of exchange. Theoretically, is is conceivable that both money functions are separated from each other. There could be one commodity serving as numêraire, and another serving as medium of exchange. However, the economies in terms of calculation costs would simply impose that both functions are wedded to a single commodity (White, 1984, p. 704).

Another advantage of a money economy consists of the desired imperfect synchronization between purchases and sales (Brunner and Meltzer, 1971). For the economic agent it can be preferable to proceed with purchases at one point in time and to undertake sales at another moment. This non-simultaneous pattern of "expenditures" and "receipts" can be made possible also in a barter economy with the help of credit. However, the barter credit refers normally to the reimbursement in terms of a specific commodity (trade posts) or, possibly, in terms of a bundle of commodities (trade intermediary). In a money economy, the "purchasing power value" of credit comprises necessarily all commodities, and credit contracts for the desired non-synchronization of receipts and expenditures are mainly replaced by money holdings. This aspect of money as a "temporary abode of purchasing power" represents the basis for the motive of the *demand* for transaction balances in contrast to the motives of the *use* of a medium of exchange.

## 2. The Economies of Scale of a Medium of Exchange

The economies in transaction costs associated with the use of money are of the internal and external type. The internal economies refer to the savings of transaction costs each indiv- idual realizes when he uses money as a unit of account and as a medium of exchange. The same is true for the use of a common language or the use of a telephone system: The communication with other individuals is easier, i.e. less costly. However, there are also externalities for others when an individual uses a common medium of exchange, a common language and a telephone system, since for the other participants the exchange of goods or informations becomes also more convenient. To the extent that the individual considers his costs of the use of money superior to his (private or internal) return, he will refrain from money transaction. If the social return of the use of money is higher than its (private or social) costs, there could be made a case for government intervention in order to generalize money trans- actions.

It should be noted that these considerations stand for the possibility of an economy which is based partly on barter and partly on money. In

general, the literature treats the alternative of an economy fully on barter terms or fully on a money standard. Consequently, the in-between case is not treated which is relevant for the gradual transition of a barter economy toward a fully monetized economy, i.e. toward complete monetary integration.

The externalities of the use of money can be illustrated by a simple model elaborated by Vaubel (1984a, pp. 33-41). For the sake of simplicity, we assume an economy consisting, for the mo- ment, of four individuals: I, II, III and IV, and where N stands for the number of individuals who are using money for transaction purposes. The individual functions of the demand for money are assumed to be identical among each other. The lowest degree of monetary integration would be an economy in which at least two individuals use money as a unit of account and medium of exchange. In Fig. 3, the individual curves of the demand for money, N = 2, are traced for the hypothesis that only two among the four individuals renounce barter and use money. The symbol m stands for real cash balances. The rate of interest, r, is given at the level r . Each demand curve reflects the marginal return of transaction balances. We assume that there is a minimum amount of real cash balances (m ) which is necessary to be held in order to regulate the purchase and sale of goods with the medium of exchange. This minimum holding of cash balances is assumed to be independent of the transaction volume.

From the individual point of view, there would be no reason to use money when monetary transactions take place only between two individuals, since the marginal return of money holdings (which are necessary for the use of money as a medium of exchange) would lie below their marginal opportunity cost. Thus, for instance, if only I and II would start the money economy, they would come to the conclusion to continue the exchanges in barter terms provided that they are not able to persuade individual III to join the money economy. The participation of III is advantageous for all three since the marginal return of real cash balances increases for each of the three from N = 2 to N = 3 (Fig. 3). The membership of III to the money economy would confer externalities – Vaubel calls them "transaction cost externalities" (1984a, p. 33) – by the amount of PR for the account of I (ST) and II (VW). Similarly, when IV joins the money economy, he creates externalities (YZ) for I (TU), II (WX) and III (QR) where YZ = TU + WX + QR.

In order to establish a completely monetized economy, the critical question concerns the means by which one could motivate III and IV to accomplish the process of monetary integration.

1. One possibility would be a once-and-for-all government subsidy of PR for III. Individuals I and II could also compensate III for the externalities they have received by him at the amount of PR = ST = VW. However, they would be net losers in the starting period because then

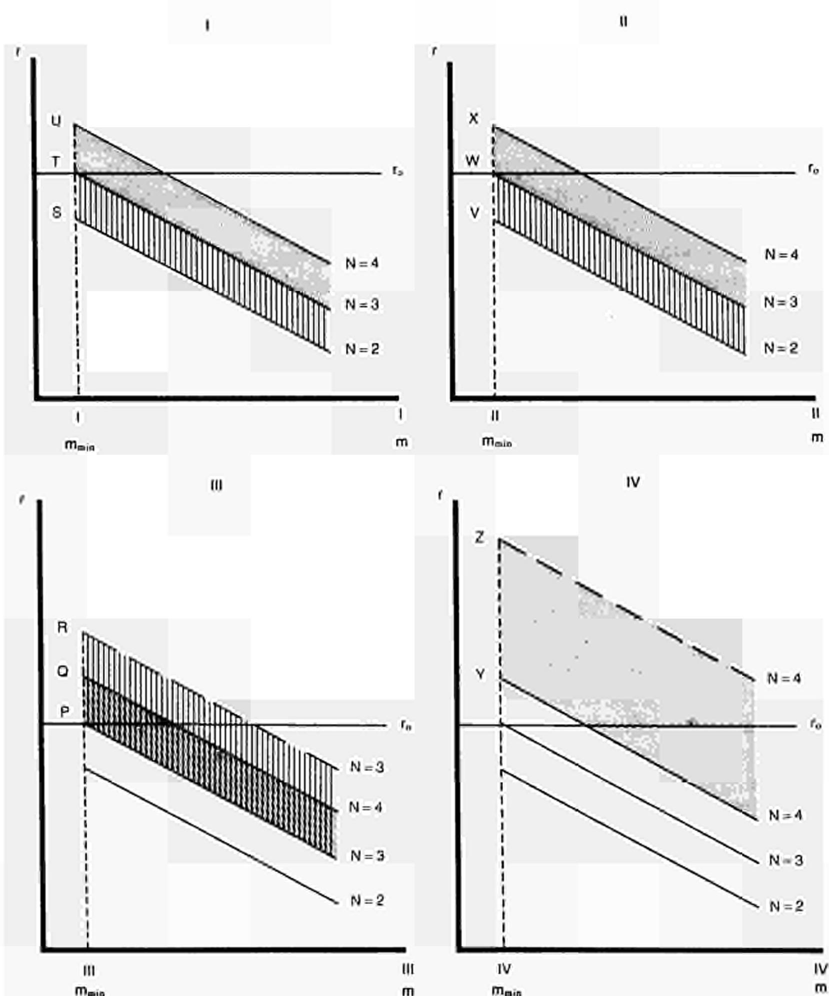



*Figure 3*

their net marginal return of money holdings is below the opportunity cost $r_0$. It should be noted that there is no necessity to subsidize individual IV since his association to monetary exchanges will be profitable for him at any rate; he could try to obtain some compensations from I, II and III for the externalities he has created for them even though they may behave as free riders. Consequently, the critical marginal money user is III or the "'critical mass' which the system must attain to be viable" (Vaubel, 1984, p. 38) is, in our example, N = 3.

2. Another possibility for the full monetization process of an economy concerns our previous argument of the invisible hand emergence of money from a barter economy in which there is an increasing number of commodities. The efficiency criterion of a money economy could motivate most, if not all individuals to use money instead of barter. Since our "critical marginal money user" could be anyone among our four individuals, all four could decide, spontaneously or rationally (and thus, without command by government subsidies), to engage in monetary transactions. This possible outcome is more likely in our model of identical demand functions for money since everybody draws the same amount of advantages from the use of money. Further, when all four individuals begin to use a medium of exchange *simultaneously*, they confer to each other the externalities so that all externalities are *mutually* internalized.

3. Until now, we have not mentioned which agency or individual supplies the money – a subject which will be treated in section IV. As we shall see, the most probable money which emerges during the transition of a barter economy toward a money economy will be a commodity money. If we choose gold as a likely candidate for the commodity money and if the suppliers are private gold mines, it will be in their profit interest to "overcome the infant industry stage" (Vaubel, 1984a, p. 41) and to subsidize the critical marginal money user or, more likely, the three money users I, II and III to the extent that I and II have not yet agreed to use gold as medium of exchange. The money producers will only supply gold coins when the cost of coinage is contained in the face value of the coin which is equal to the gold content plus the cost of coinage, plus a seigniorage profit in case of monopoly. If the cost of coinage underlies decreasing production costs, the money producer(s) will subsidize III, or I, II and III for the start-up costs which will be financed by future profits resulting from the use of money by IV (and others).

## 3. The Economies of the Use of a Single Medium of Exchange

The function of money as a means of payment refers to a *generally acceptable* means of payment. The general acceptability may differ between areas and between types of transactions. The latter case concerns the "technical" question according to which different kinds of the same currency unit (coins, bank notes, sight deposits) are not complete substitutes such that for many cases, each of them can be used more easily for certain types of payments. However, the real economic issue is related to the area of the money circulation. If there are many monies within an area, for instance, within the largest one, which is the world economy, the optimum number of monies will be one and this for the following reasons (Claassen, 1984, pp. 51-53).

1. As Swoboda (1968) has shown, the average holding of transaction balances will be more important, the higher the number of monies. This result is based on the assumption that certain payments (for instance, intra-national ones) can only be effected in one money and other types of payments (for instance, international ones) in another money.[3]

The necessary use of different monies (which is analogous to the above "technical" question of the existence of different kinds of the same currency unit) may not be the result of legal restrictions with respect to the exclusive use of a certain money. It could (but must not) be conceived equally in a world of private money producers without any government intervention.

2. Calculation costs and currency conversion costs are another item which reduces the optimum number of monies to one. Calculation costs are minimized when there are fixed exchange rates and when monies are exchanged in a relation of 1:1. In this case, the calculation costs are approximately zero because such a fixed-exchange-rate system resembles a one-money world. But even in such a world currency conversion costs are still existent due to the conversion of one money into another when payments have to be made in this other money. One reflection of these currency conversion costs are the opportunity costs of the average higher transaction balances individuals must hold and which we have already mentioned above.

As Vaubel (1984a, pp. 41-42) shows, the economies of currency conversion costs have also an internal and external component. An individual who used formerly currencies X and Y and chooses now only currency X, will have lower – internal – transaction costs in terms of the savings of lower transaction balances and of lower conversion costs. He will also confer externalities on the other users of currency X (and he will withhold external benefits from the users of currency Y).

To the extent that the money supply is competitive producing several currencies linked to each other by flexible exchange rates, it is conceivable that the competitive market forces drive the private money producers to currencies with, successively, stable purchasing power, fixed exchange rates and, eventually, with a denomination in a relation of 1:1. This possibility has been envisaged by Hayek (1984, p. 39):

> I believe that complete freedom to offer to the public alternative monies would rapidly lead to a number of types of money, all of them essentially stable in value, all widely known for their quality and – this is perhaps a surprising feature – all of them stable in terms of each other. They would represent more or less the same store of value under different names . . . These monies, I believe, would be partly expressed in denominations of the same magnitude, though bearing different names according to the issuer of a particular type of token or money. It is a strange picture, I admit, but the more one thinks about it, the more realisable it appears.

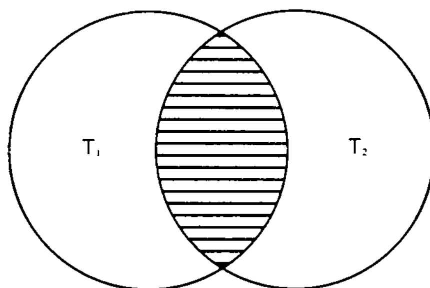



*Figure 4*

The monetary quality of a single, universally accepted means of payment is the highest one in comparison with the existence of several monies. For instance, the existence of two monies can be compared with the existence of two telephone systems, $T_1$ and $T_2$ in Fig. 4. In the non-shaded section of the area $T_1$ ($T_2$), all telephone users can communicate with each other via the system $T_1$ ($T_2$); the monetary analogy would be that in this area section the only money used is the money of type $T_1$ ($T_2$). In the shaded section, both systems are used. One possible (and the most costly) type of organization would be that their inhabitants must possess a second phone receiver. In order to eliminate the "calculation costs" the phone number of each user should be the same in both telephone systems. Thus, the remaining costs – the "conversion costs" – consist of "holding" two receivers for the inhabitants of the shaded section. At first sight, these transaction costs may not be considerable, but two further considerations make them more relevant.

Suppose that each area is penetrated completely by both tele- phone systems (the shaded surface of Fig. 4 would cover both circles). To merge both systems into a single one which, from a technical point of view, could imply a simple, nearly costless linking of the lines between the two telephone centres, would (or could, depending on the production function) create economies of scale of which the telephone users profit by having only one single phone receiver and by paying lower charges due to the lower average production costs. The lower price implies that the natural monopoly is regulated in the sense that the price of phone calls is equal to the average cost (or, in the case of the natural monopoly of money creation, that the seigniorage gain is transferred to the money holder). If this condition is not ful- filled, then it may be that the former oligopolistic structure was less costly for the telephone users even though, under the aspect of the whole economy (producers and users), more resources are tied up in the telephone industry.

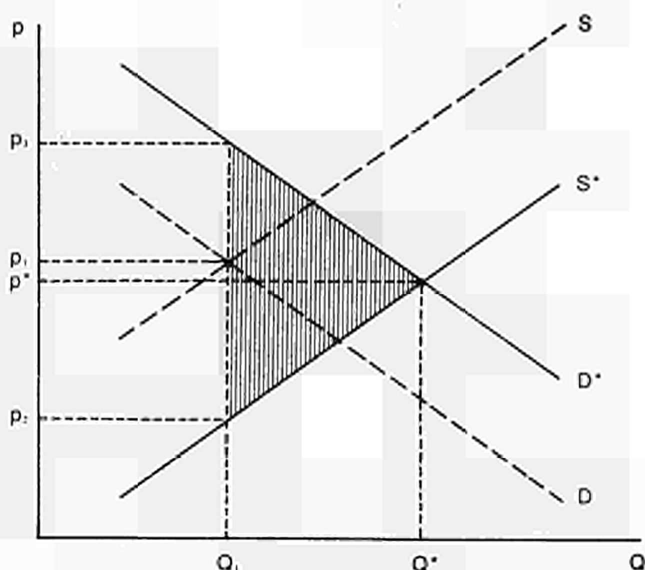



*Figure 5*

2. Instead of two telephone systems, suppose there exists a competitive system of n telephone companies covering simultane- ously all areas. The economies of merging several companies will be extremely high. On the one hand, the telephone users will welcome such a monopolization because of their lower transaction costs provided – in the absence of any monopoly regulation – that the price of the product is not increased. On the other hand, among the producers, there will be a tendency to merge the production such that, in the end, an oligopolistic or monopolistic structure of production will develop. Both aspects could explain the regional monopolization of money production which would have ended in a world monopolization in the absence of government monopoly *and* by neglecting the second quality of money which is its store-of-value aspect.

Before analysing the realization of this second quality of money through a monopolistic or competitive money supply process, we summarize our results with respect to the economies of transaction costs when the concerned (national or world) economy converges gradually from a (non-organized or organized) barter economy to a money economy in which the unit of account and the medium of exchange reach eventually their highest degree of moneyness in terms of a single medium of exchange – a situation which we have called perfect monetary integration.

Lower transaction costs imply lower prices for goods and a higher production volume. In Fig. 5, the demand and supply conditions of a good

are represented with and without transaction costs. In the case of the non-existence of transaction costs. the demand and supply functions are $D^*$ and $S^*$, respectively, the equilibrium price $p^*$ and the equilibrium output $Q^*$. Transaction costs increase the supply price (production cost $S^*$ plus transaction costs of the suppliers) and decrease the demand price (by the transaction costs of the buyers). In Fig. 5, transaction costs are considered as an excise tax on the supplied and demanded quantities. The equilibrium output is reduced to $Q_1$. The equilibrium price is $P_1$ to the extent that it comprises the transaction costs. The "pure" production price is $P_2$ and the "gross" demand price is $P_3$.

The shaded triangle in Fig. 5 illustrates the economies of transaction costs when the economy converges, e.g., from a barter economy to a money economy with several media of exchange and from that to an economy with a single medium of exchange. A more efficient form of exchange organization (resulting, respectively, from non-organized barter to organized barter, from an organized barter to a money economy with increasing degrees of monetary integration) changes necessarily relative prices and the volume of exchange.[4]

## IV. Monetary Stability: A Monopolistic versus Competitive Framework for the Money Supply

While the last section emphasized the demand side of money since it was concerned with the private and social *use of money*, the present section treats the supply side of money in terms of an appropriate institutional framework for the *production of money* in order to guarantee price stability.

Economies of scale *exist* with the use of money and they *can* also exist with the production of money. If the second proposition reveals to be relevant, there will be a case for a monopolistic structure of the money supply. The latter would assure monetary integration in terms of a single medium of exchange. However, if a monopolistic production does not realize monetary stability, the two fundamental objectives of the monetary constitution would be antagonistic to each other within a monopolistic framework of the money supply.

If one evokes the monopoly of the money supply, it is not necessarily a state monopoly, but it can also be a private one. Government intervention could be justified with the existence of externalities. As far as the use of money was concerned, we had come to the conclusion that there were economies of scale – of the internal and external type – associated with a more gener- alized use of a common medium of exchange. However, as we have shown, the externalities could be appropriated mutually by the

members of the money economy so that there would be a sufficient incentive for them to start off with monetary transactions without any intervention by government. Consequently, from the point of view of the economies of transaction costs connected with the use of a (single) medium of exchange, the supply of money could be provided by the private sector of the economy. Whether the (private) supply should be monopolistic or competitive is another question which will be discussed under the following three aspects. On the one hand, price stablity may be better maintained by a competitive system. On the other hand, internal economies of scale in the production of money and, in particular, in the producation of monetary confidence may imply the existence of a natural monopoly even though the latter can be run by a private producer; in order to avoid the possibility of monetary instability, free entry conditions into the money industry should be maintained. Finally, a government role in monetary affairs is often proclaimed with another aspect of externalities which concerns the possible financial breakdown of the monetary system. Whether "liquidity crises" could be avoided efficiently by the market mechanism replacing the traditional government role of a lender of last resort will be analysed at the end of section IV.

## 1. Monopolistic versus Competitive Supply of Money and the Objective of Price Stability

In the "older" literature, the monopolistic constitution of currency creation (base money) by government is supported by the argument of a necessary condition for price stability (Vaubel, 1978, pp. 59-60):

> . . . it has not the slightest sense to render the money creation to the free competition of the private sector . . . Since, in principle, each quantity of money can buy all goods provided that prices are adjusted, the desire to possess possibly the highest amount of it is senseless. A higher quantity of money leads only to a higher price level . . . Money creation belongs to the activity of government. (Lutz, 1936, pp. 4-5.)
>
> So long as the fiduciary currency has a market value greater than its cost of production – which under favorable conditions can be compressed close to the cost of paper on which it is printed – any indivudual issuer has an incentive to issue additional amounts. A fiduciary currency would thus probably tend through increased issue to degenerate into a commodity currency – into a literal paper standard – there being no stable equilibrium price level short of that at which the money value of currency is no greater than that of the paper it contains. And in view of the negligible cost of adding zeros, it is not clear that there is any finite price level for which this is the case. (Friedman, 1959, p. 7.)
>
> Competition is . . . inappropriate for determining the amount of fiat currency. (Friedman, 1951, p. 211.)
>
> The banking industry will destroy itself as a money producer by being competitively driven to deteriorate the quality of the product. (Pesek and Saving, 1967, p. 129.)

> A competitive banking system would be under constant incentive to expand the nominal money supply and thereby initiate price inflation . . . Stability in the trend of prices (a special case of which is price stability) and in the trend of expectations about the future course of prices – which are generally agreed to be important to the social welfare – requires social control over the total quantity of money supplied by the banking system. (Johnson, 1968, p. 976.)
>
> Control of base money remains the responsibility of the government or of a chartered monopolist, and this monopoly must continue to assure that society retains the benefit of producing money at lowest cost. (Meltzer, 1969, p. 35.)

The Hayek proposal of 1976 pretends just the opposite of the traditional view on the state monopoly of currency creation. Since history shows us monetary periods of high monetary instability associated with government monopoly of the money supply, he conceives a competitive money supply process as a producer of a better quality of money in terms of higher monetary stability:

> Governments prevented the people from searching for the right solutions and experimenting with a process of selection which would have led to continuous improvements. The monopoly prevented a spontaneous formation of money in ways similar to what we have seen in the area of law, language and morals, where, through a process of evolution, the more effective forms displaced the less effective forms. If we had been allowed to benefit from a similar form of selective evolution where money was concerned, we would have had a money which would have been entirely different from the money we have today. (Hayek, 1984, p. 31.)

Recent literature has shown that a system of privately issued and competing monies assures price stability. One of the most original writers in this field is Klein (1974). He has shown that, generally, the case of price instability in terms of an infinite price level can be made when money producers offer an indistinguishable homogeneous money product (for instance, by not placing their names on the money they issue) – a situation which furthermore implies fixed exchange rates among the various money products. In such a case, any money producer can overissue money which is indistinguishable from other monies such that competition would lead to an infinite price level. A competitive interest payment on money would not prevent the overissue because the interest rate on money would reflect the higher inflation rate (even by supposing that it is expected correctly) where the latter one can have whatever level. This special case was probably in the mind of the authors we mentioned above, who postulated the government regulation of the money supply.

If each money producer is enforced to "print" his particular name on his money product, counterfeiting by overissue becomes a less relevant case. Now the (interest-bearing) money products are differentiated among each other – a situation which involves flexible exchange rates (unless there is a government or privately issued dominant money – (high)-

powered money – into which the other monies are convertible). A profit via an overissue by a single firm can be realized to the extent that some part of the inflation rate, in terms of the particular money, is not anticipated. This overissue implies a disinvestment in the money producer's "brand-name capital" (Klein, 1974) – the confidence in the money concerned will sharply decrease with espect to other monies, such that the firm will be driven out of the market. Consequently, the money producer's decision to overissue will be determined by the comparison between the short-run profit of over-production and the long-run profit yielded by staying in the market.[5]

Consequently, a competitive system of privately issued monies is viable with respect to the maintenance of a stable price level. It is even conceivable, as Hayek emphasizes (1984, p. 39) and as we have mentioned in section III (p. 21), that the competition mechanism among distinguishable monies (i.e. monies with different names) with stable purchasing power leads to denominations of the same magnitude – in the very extreme case to an exchange relation of 1:1 – such that a perfect monetary integration could also be realized in the course of time (even though the monies remaining distinguishable from each other are linked to each other on the basis of flexible exchange rates where the latter would be stable).

However, a case against currency competition could be made, not on the grounds of monetary stability which can be refuted as we have seen, but along the lines of the existence of economies of scale in the production of monetary confidence, i.e. in the production of information about the quality of a stable money. This case concerns the natural monopoly hypothesis of the supply of money.

## 2. The Natural Monopoly Hypothesis of the Money Supply and the Creation of Monetary Confidence

In a situation of "natural" monopoly, economies of scale are so important that one firm can always produce more cheaply than can any large number. Public utilities such as electricity, telephone, gas are commonly believed to belong to this category. Traditional economic policy with respect to natural monopoly has been to retain the single large firm but to prevent monopolistic exploitation through regulation.

Does money production constitute a natural monopoly? The answer is affirmative if one can show that there are production costs of money and that the average production costs are falling with higher output (Claassen, 1984, pp. 50-54).

With respect to the exisstence of the production cost of money, one has to distinguish between the cost of producing the quantity of money and the cost of producing the monetary services which are associated with

the quantity of money. The costs of producing the quantity of money may be extremely low: In the case of paper money produced by a central bank they refer to the cost involved by the employment of a printing (and bookkeeping) machine plus the workers ("officials") who serve the machine, sell the money and administer the money circulation; the production costs of sight deposits by banks imply more bookkeeping, but they are essentially of the same nature as the former ones. To the extent that these costs are mainly fixed costs *and* that the average production cost is falling with a higher quantity of money, a (potential or effective) natural monopoly may emerge which operates within a certain region, which coincides with the national borders, is extended to a larger area or covers the whole world, depending on the final limit of the economies of scale. The regulation of such a monopoly should be interpreted as aiming at making the economic profit equal to zero (i.e. that seigniorage gain which lies above the normal profit necessary to attract and retain the resources employed in the money industry) in order to obtain the optimum quantity of money.

However, this view is a rather naive one because it neglects the quality aspect of the money product, i.e. its monetary services. There are costs for producing a certain *quantity* of money and there are *other* costs for producing a certain *quality* of money where the quality refers to the degree of moneyness of the money products. Traditionally, this degree of moneyness refers to the two functions of the quantity of money: to money as a means of payments and to money as a store of value.

Money is generally accepted as a means of payment when people have confidence in this money and this confidence is basically founded on its purchasing-power-value stability. Confidence belongs to the domain of information and information is based on past experience and future expectations. Is follows that newly created monies and existing monies imply completely different degrees of confidence. Thus, for instance, a newly created money which lacks, by definition, any historical dimension of (good or/and bad) value-stability behaviour has to be equipped by resources which create confidence, and these resources may consist of backing the money by commodities, by other currencies as reserves, by respectable names in bank management or by various other techniques. Consequently, the money producer has costs of selling (and not only producing) the money which are equivalent to the investment cost of building up a "brand-name capital", as Klein (1974) calls it, which assures a certain degree of monetary confidence.

Under this aspect, certain past monetary experiences no longer constitute a puzzle for monetary theory. If one looks at the gold standard, gold was not necessarily a costly money for the economy compared with paper money, because the cost of constituting the "rand-name capital" in terms of the gold content of money can be considered to be lower at that time

than the investment cost of creating confidence associated with paper money.[6]

Friedman and Schwartz (1985, p. 13) indicate that monetary history does not provide any single example where money emerged on the basis of an abstract unit of account – our present fiat unit. In the starting period, all monies had been linked to a commodity. Even our present monetary systems had been founded on a commodity standard over centuries and, as we have mentioned in the introduction, it is only very recently – i.e. since the beginning of the 1970s – that their last remaining relics of a commodity money had been removed and that they are only now on a pure fiat standard.[7]

Consequently, the exclusive emergence and long-lasting life of commodity monies are explained by the necessity for creating and maintaining monetary confidence. These lessons from monetary history are not only valid for *government money*, but also for the production of *private moneys*. The two standard examples of private competitive note issue in monetary history is that of Scotland (1749 – 1845) and of various parts of North America during the 19th century (Vaubel, 1984b). As Friedman and Schwartz (1985, p. 14) have emphasized with respect to these monetary experiments of competition among private money producers, there cannot be found any single case of a privately produced "pure fiduciary" currency:

> Historically, producers of money have established confidence by promising convertibility into some dominant money, generally, specie. Many examples can be cited of fairly long-continued and successful producers of private moneys convertible into specie. We do not know, however, of any example of the private production of a purely inconvertible fiduciary moneys (except as temporary expedients, e.g., wooden nickels, clearing house certificates). . .

From these historical monetary lessons of private money production one could derive the conclusion that, at the present, any private money would only have a chance to be accepted when it is on a "commodity standard". And to cite again Friedman and Schwartz (1985, p. 19):

> Historical experience suggests that the only plausible alternative to a government issued fiduciary currency is a commodity currency, with private issuers producing inside money convertible into the commodity. And we believe that even that outcome is highly unlikely unless there is a major collapse of national currencies – something approximating hyperinflation on a worldwide scale.

When the high investment costs for creating and maintaining monetary confidence (with respect to public money as well as with respect to private money) are coupled with decreasing costs of disseminating information about the quality of a money, then the possibility – but not the necessity – for the emergence of a natural monopoly can arise. There could be made a case where "the cost of disseminating information about a money

declines, as the number of issues decreases." (Vaubel, 1984a, p. 46.) If it is true that the production of the store-of-value service of money follows the principle of decreasing cost, then the existence of a natural monopoly would be justified. However, on the one hand, our present knowledge about the cost structure of a money producer is still insufficient. On the other hand, if a natural monopoly reveals to be the true solution, there could be made no case that government should be the money producer.

As far as the first point is concerned – our imperfect knowledge about the cost behaviour of the production of monetary confidence, the monetary constitution should always guarantee the free entry of any potential money producer. As Vaubel (1984a, p. 47) puts it:

> Since . . . we cannot even be sure that money or the currency unit is a natural monopoly, the case against restrictions of entry is overwhelming. Only if a governmental producer of money can prevail in conditions of free entry and without discriminatory subsidies, he is an efficient natural monopolist.

But even if a natural monopoly is the natural issue, there is no necessity that it should be held by government. It can be run also by a regulated private monopoly. The reasons for government monopoly are various, but fundamentally fallacious:

– because of stabilization policies (government can pursue them also in terms of its own money beside the existence of other private monies when there is no clear-cut case for a natural monopoly);

– because of the "legal-tender" character of government money (but legal tender is not a sufficient condition to create monetary confidence);

– because of the constitutional existence of an independent central bank whose obligation is the maintenance of price stability (officials do not own the capital of the central bank and therefore they have less incentive to conserve it than a private producer);

– because of government as a "lender of last resort". To this issue we shall turn now.

## 3. The Government's Role as a Lender of Last Resort

Welfare economics have taught us that government intervention may be justified or even a necessity if there is some market failure. One type of market failure consists of the phenomenon of externalities. The stability of the financial system and its maintenance could be conceived as producing positive externalities, and their contrary, the instability of the financial system and its possible collapse, as a case of negative externalities (Solow, 1982). The financial instability can be prevented by the central bank in its function as a lender of last resort (LLR) and by an insurance agency in insuring deposits against bank failures. However, both institutions create

the phenomenon of moral hazard in terms of more risk-taking by banks. If credit surveillance by the central bank or by the insurance agency is insufficient, another solution called market solution could be envisaged according to which it is in the proper self-interest of banks to produce sufficient funds to other banks being exposed to a temporary liquidity problem. (Claassen, 1985, pp. 219-224.)

*Externalities*. The instability of the financial system in present monetary systems is caused by the appearance of a lack of confidence depositors of a bank (and of any other financial institution) may have with respect to the bank's ability to convert the deposits – sight or time deposits – into cash. A resulting run on the bank would involve its default provided it cannot sell immediately its assets at a "reasonable" price or borrow sufficient external funds (liquidity problem). The default by one bank can induce the default of other banks which constitutes the proper phenomenon of negative externalities. That other banks can also be confronted simultaneously by a "convertibility test" (bank run) results from a subsequent generalization of lack of confidence by the public in other banks or at leat in those banks which have credit relationships with the first bank on which the run began. Consequently, a default by many banks is a conceivable outcome and depending on their relative size within the financial system and on the duration of the default, it may be identified with a collapse of the financial system.

In most, if not all cases (when one excludes situations of wars or catastrophes), the very beginning of the lack of confidence is based on the default of some debtor A who is not a bank, but an enterprise and who is indebted towards B and where B in his turn may also be debtor against C. Thus, a default of A may imply the default of B and C. This chain reaction creates an externality for the lenders of C to the extent that they are well able to judge the creditworthiness of C but not the soundness of the whole network of interrelated debtor-creditor relations. If the market does not provide a sufficient transparence of the interlocking debt structure, government should intervene by forcing such a transparence and by imposing credit standards. However, the nature and size of externalities are fundamentally different if B and C are commercial banks operating on the basis of a fractional reserve requirement system, since then the specific confidence problem arises and the subsequent run on banks is set in motion with the consequences of the instability threat to the whole banking system and of its repercussions on the real sector of the economy.

*Central Bank as LLR*. How could government interventions avoid this type of externalities? One possible way would be to introduce a 100 per cent reserve requirement system, at least for sight deposits, which would exclude any lack of confidence by depositors and, by this, any bank run in terms of a convertibility test from sight deposits into currency.

If one remains within our present monetary systems of fractional

reserve requirements, the other solution would consist of assigning to the central banks the function of the LLR. The role of central banks for dealing with liquidity crises has already been discussed in the literature of the 19th century and favoured by Thornton (1802) and Bagehot (1873). The essence of this discussion was that only illiquid and not insolvent banks should have unlimited access to the central bank. Thus, an insolvent bank should not be preserved from failure, but the possible cumulative chain reaction in terms of runs on other banks – the proper externality aspect – should be avoided by the LLR-actions of the central bank. Thus, Bagehot argued that only "fundamentally sound" banks should be rescued and that the Bank of England should lend unlimitedly, but at a penalty rate, since only a fundamentally sound bank would be able to pay off the high interest rate.

*Deposit insurance.* Another solution for avoiding any bank run is the adherence of the banks to a deposit insurance. Such an insurance system would wipe out the appearance of a bank run only to the extent that *all* deposits are insured.

In the present compulsory deposit insurance systems existing in nearly all industrialized countries, the insurance is mostly limited to an upper amount of a bank client's deposit (e.g. in the USA to * 100,000). The motivation for introducing a compulsory deposit insurance was initially the paternalistic protection of small depositors assuming that large depositors have sufficient information about the quality of the bank assets. However, if the main function of a deposit insurance system is conceived as that of the prevention of a run on banks, then each deposit amount should be insured.

According to Friedman (1959, p. 38), the Federal Deposit Insurance Act has made the LLR function of the Federal Reserve System obsolete for the USA since depositors will not engage in any sudden wide-spread demand for currency at their banks anymore, ". . . not because the LLR function has been taken over by someone else but because it no longer needs to be performed". This optimistic view only holds if the deposit insurance system is of the generalized (comprehensive and complete) type. On the other hand, at least at the very beginning of the insurance system, the need for the LLR function may still exist to the extent that the insurance company has not accumulated a sufficient amount of reserves out of the insurance premiums over the past.[8]

*Moral hazard.* The prevention of liquidity crises by the LLR function or by a generalized deposit insurance system raises a well-known problem of moral hazard – a phenomenon which is omnipresent in all types of insurance contracts and according to which the insured individual is less careful in preventing damages than in the case of the inexistence of the insurance. The example of a fire insurance is often cited in the literature (Solow, 1982, p. 243). The social benefit of fire insurance consists of the

reduction of uncertainty among risk avertors. Its social cost refers to the increased number of fires which occur if the insured are less careful about fire prevention. There is a net social advantage from fire insurance to the extent that the lower individual cost of risk-taking outweighs the additional incidence of damages from fires.

The analogy of the net social benefits arising from the existence of a LLR or of a deposit insurance with those of a fire insurance is rather doubtful since the extra damages of a more risk-taking bank may be disastrous for the whole economy compared with the extra fire damages of the less careful house proprietors or tenants. More risk-taking by banks increases the probability of insolvency and, by this, the probability of the instability of the financial system. The comparison with the fire insurance concerning the social cost of moral hazard would be more adequate if the increase of fire damages implied the danger for wide areas of towns to be destrozed by fire which is nothing else but the manifestation of the externality aspect.

The existence of moral hazard suggests that insurance should be partial (coincurance) involving that the central bank in its function as a LLR offers only a limited amount of currency or that the deposit insurance indemnifies only one part of the outstanding deposit volume.

However, such a partial insurance scheme would not exclude the possiblity of bank runs and of a financial collapse. Consequently, insurance coverage should be complete if financial instability is to be avoided. But then moral hazard would again be omnipresent. It could be reduced either by randomizing the action of the LLR or of the deposit insurance, or by imposing strict creditworthiness standards. Because the first solution does not solve the confidence problem of depositors, the second one is mostly chosen.

*Market solution*. For several reasons one could doubt the usefulness or the necessity of the LLR function or of the compulsory deposit insurance if the market already fulfils the function of "internalizing (or avoiding) the externalities". A bank, being "fundamentally sound" but facing a liquidity problem due to a temporary lack of confidence of its depositors, may obtain the necessary funds by borrowing from other banks. To the extent that it cannot raise a sufficient amount, it may offer a higher price as it would also have obliged to do in the case proposed by Bagehot (1873) where the Bank of England should require a penalty rate.

The counter-argument to the private-market solution of a liquidity crisis consists of indicating the imperfections of the credit markets. Since information about the future default risk of any bank debtor is imperfect, credit may be granted to a limited amount even with collateral lending. According to Stiglitz and Weiss (1981), an excess demand in the credit market may not produce a rise in the interest rate since it drives from the credit market those borrowers with low default risks and attracts those

with high default risks. This adverse selection process driving the low-risk borrowers out of the market would imply that the lenders are stuck with the "lemons". The consequence of a higher interest rate could be a decrease in the expected return on banks' assets such that banks would practise credit rationing by trying to satisfy the credit demand of low-risk classes. Others in high-risk classes including a bank facing a liquidity crisis would obtain only a limited amount of credit or nothing at all. This dilemma for the above credit-seeking bank has already been noticed by Bagehot (1873, p. 69): "Every banker knows that if he has to *prove* that he is worthy of credit, however good may be his argument, in fact his credit is gone".

Despite the *prima facie* convincing argument according to which a sound but illiquid bank may not have a sufficient access to the credit market, another consideration may render it completely obsolete. Banks are conscious of the externality aspect as the consequence of the illiquidity of a bank and, in the case of an institutional and factual inexistence of any LLR function by the central bank, it is in their own interest to rescue the illiquid bank since otherwise they may suffer the same fate. Furthermore, the information concerning the assessment of whether the concerned bank is either illiquid or insolvent is not *a priori* more detailed and accurate for the central bank than for the private banks since otherwise there would exist a certain priority to the LLR function of the central bank. From this point of view, Vera Smith (1936, p. 148, cited by Friedman and Schwartz, 1985, p. 27) concludes that:

> A central bank is not a natural product of banking development. It is imposed from outside or comes into being as a result of Government favours.

But there are two cases where one could doubt the efficiency of the solution of the liquidity crisis by commercial banks. First, the illiquid bank may be so big that the other banks are not capable of collecting a sufficient amount of currency without the aid of the central bank. Secondly, a certain number of commercial banks may behave as free riders, in particular when there is some doubt about the "soundness" of the illiquid bank so that the other banks may not procure the necessary volume of currency.[9]

Consequently, the market solution does not represent a clear-cut superior alternative for *all* circumstances of bank illiquidity compared with the government solution. On the other hand, the government solution implies social costs in terms of moral hazard since credit surveillance by the central bank or the deposit insurance agency does not absolutely exclude more risk-taking by banks. Thus, the disadvantage of one solution (insufficient rescue measures by banks in the illiquidity case of a single large bank) has to be weighted against the disadvantage of the other (moral hazard).

The above described market solution could be conceived to function

better if banks create their proper private deposit insurance system since the stability of the financial system and the avoidance of instability by the creation of a private insurance agency are in their own self-interest. If some banks do not subscribe to the deposit insurance, they may behave like free riders, but then either their depositors will shift to the insured banks or the non-insured banks have to pay a higher interest rate on deposits as an equivalent to the default risk premium. Since the phenomenon of moral hazard is also existent for a private deposit insurance system, partial insurance coverage may be the appropriate means of limiting it. A run by the public on *illiquid* banks is cushioned, if not even excluded by the confidence of the public into the rescue operations by other banks and/or by the private insurance agency operating as a LLR.[10]

## V. Concluding Remarks

If one looks at monetary history which presents a large range of monetary experiments, one could emphasize the fact that there have been waves of monetary integration followed by monetary dis-integration (several moneys, flexible exchange rates). The latter one always emerged with increasing monetary instability. Our contention is that there is no monetary constitution which guarantees *permanently* the *optimum optimorum* in terms of monetary integration and monetary stability (and of the optimum quantity of money). The arguments in our paper have shown that the targets of monetary integration and monetary stability may be, at one time, opposing forces and, at another time, converging forces.

There is, therefore, no *a priori* first-best solution for the monetary constitution. But this conclusion does not mean that the government monopoly of money production is superior. We hope that the arguments in this paper have shown that a case can be made that a privately run production of money (competitive or monopolistic, but with existence of potential competition) is superior to government monopoly. However, this private production of money does not imply, for the whole future, the highest degree of monetary integration and the highest degree of monetary stability which are the fundamental aims of each monetary constitution.

As we have stressed in the introduction, the recent discussion of alternative (i.e. private) monetary systems was motivated by the highly inflationary bias of the world fiat standard during the 1970s. However, since 1980, we observe a deceleration of the world inflation rate and there is no inherent reason why our present monopolistic fiat standard must necessarily produce again higher monetary instability in terms of waves of accelerated and decelerated inflation rates. This optimistic view is based eventually on the *political motivation* for inflation which has been attenuated progressively over the last years for several reasons.

On the one hand, the general belief in any (even short-run) trade-off

relationship between inflation and unemployment has collapsed with respect to past and recent experience, and it could be argued that just the opposite – a positively sloped Phillips curve – is relevant for the present time. However, such a belief does not exclude the possibility that in some future, after having attained monetary stability, there will be again a revival of a presumably existing (at least short-run) trade-off relationship. On the other hand, government revenues from inflation which could have been initially the cause of inflation, have declined considerably. The first government revenue resource from inflation is the paper money issues themselves (the "inflation tax" on real holdings of high-powered money); however, this "hidden" tax provides a low government receipt of less than 1% to the extent that the inflation rate is moderate (e.g. 10%). The second component of government revenue from inflation is that of a partial indexation of the personal income tax; however, political pressure during the inflation process leads to a more or less full indexation scheme. Thirdly, the reduction in the real value of outstanding government debt issued at interest rates that did not take into acccount sufficiently the allowance for future inflation, has been sharply eroded by the developments in financial markets which have become increasingly conscious ("rational") of the future evolution of inflation.[11]

## References


Alchian, A.A., "Why Money?", *Journal of Money, Credit and Banking*, 1977.

Bagehot, W., *Lombard Street*, London 1873, reprinted Homewood, Ill., 1962.

Barro, R.J. and Fischer, S., "Recent Developments in Monetary Theory", *Journal of Monetary Economics*, 1976.

Brunner, K. and Meltzer, A.H., "The Uses of Money in a Theory of an Exchange Economy", *American Economic Review*, 1971.

Claassen, E., "Monetary Integration and Monetary Stability: The Economic Criteria of the Monetary Constitution", in: P. Salin (ed.), *Currency Competition and Monetary Union*, The Hague, 1984.

Claassen, E., "The Lender-of-Last-Resort Function in the Context of National and International Financial Crises", *Weltwirtschaftliches Archiv*, 1985.

Dornbusch, R., Stopping Huperinflation: Lessons from the German Inflation Experience of the 1920s, Working Paper No. 1675, National Bureau of Economic Research, 1985.

Friedman, M., "Commodity-Reserve Currency", *Journal of Political Economy*, 1951.

Friedman, M., *A Program for Monetary Stability*, New York, 1959.

Friedman, M., Monetary Policy in a Fiat World, Hoover Institute, mimeo, May 1985.

Friedman, M. and Schwartz, A.J., Has Government any Role in Money?, Hoover Institute, mimeo, June 1985.

Hayek, F.A., *Choice in Currency: A Way to Stop Inflation*, London, 1976a.

Hayek, F.A., *Denationalisation of Money*, London, 1976b.

Hayek, F.A., "The Future Unit of Value", in: P. Salin (ed.), *Currency Competition and Monetary Union*, The Hague, 1984.

Johnson, H.G., "Problems of Efficiency in Monetary Management", *Journal of Political Economy*, 1968.

Klein, B., "The Competitive Money Supply", *Journal of Money, Credit and Banking*, 1974.
Lutz, F.A., *Das Grundproblem der Geldverfassung*, Stuttgart, 1936.
Meltzer, A.H., "Money, Intermediation and Growth", *Journal of Economic Literature*, 1969.
Menger, C., "On the Origin of Money", *Economic Journal*, 1982.
Pesek, B.P. and Saving, T.R., *Money, Wealth and Economic Theory*, London, 1967.
Sachs, J.D., *Theoretical Issues in International Borrowing*, Princeton Studies in International Finance, 1984.
Samuelson, P.A., "What Classical and Neoclassical Monetary Theory Really Was", *Canadian Journal of Economics*, 1968.
Smith, V.C., *The Rationale of Central Banking*, London, 1936.
Solow, R.M., "On the Lender-of-Last-Resort", in: C.P. Kindleberger and J.P. Laffargue (eds.), *Financial Crises*, Cambridge, 1982.
Starr, R.M., "Exchange in Barter and Money Economies", *Quarterly Journal of Economics*, 1972.
Stiglitz, J.E. and Weiss, A., "Credit Rationing in Markets with Imperfect Information", *American Economic Review*, 1981.
Swoboda, A.K., *The Euro-Dollar Market: An Interpretation*, Essays in international Finance, Princeton, 1968.
Thornton, H., *An Enquiry into the Nature and Effects of the Paper Credit of Great Britain*, London, 1802; reprinted London, 1962.
Vaubel, R., *Strategies for Currency Unification*, Tuebingen, 1978.
Vaubel, R., "The Government's Money Monopoly: Externalities or Natural Monopoly?", *Kyklos*, 1984a.
Vaubel, R., "The History of Currency Competition", in P. Salin (ed.), *Currency Competition and Monetary Union*, The Hague, 1984b.
Wallace, N., "An Approach to the Study of Money and Nonmoney Exchange Structures", *Journal of Money, Credit and Banking*, 1972.
White, L.H., "Competitive Payment Systems and the Unit of Account", *American Economic Review*, 1984.

## Notes

[1] F.A. Lutz, *Das Grundproblem der Geldverfassung*, Stuttgart 1936. During the same year, another fundamental monograph on the institutional structure of the monetary system was published by Vera C. Smith: *The Rationale of Central Banking*, London 1936, to which we shall also refer. Some years later, Vera Smith became Vera Lutz.

[2] As a matter of course the zero degree of monetary integration would be a barter economy. Under this aspect a world economy with n currencies having n (n-1) flexible exchange rates can be conceived to be in the neighbourhood of a barter economy as far as the international transactions are concerned.

[3] The Baumol-Tobin formula for the transactions demand for money (M) is:

$$M = \sqrt{\frac{tT}{2r}} \quad (1)$$

where T is the real value of transactions (payments), t the fixed costs per transaction "selling bonds – buying money" and r the interest rate. Suppose the existence of n monies expressed by the same numéraire. The transactions demand for the n monies will be:

$$\sum_{i=1}^{n} M_i = n \sqrt{\frac{tT_i}{2r}} \quad (2)$$

where $T_i$ is the transaction volume in i's money and where t and r are supposed to be identical for all monies. Assume for the sake of simplicity that there is the same transaction volume in each money ($T_1 = T_2 = \ldots = T_n$) such that the total transaction volume is $T = n T_i$. Then, formula (1) can also be written as:

$$M = \sqrt{\frac{tnT_i}{2r}} \tag{1a}$$

Consequently, the transaction balances in a world of n currencies are higher than in a single money world by:

$$\sum_{i=1}^{n} M_i - M = (n - \sqrt{n}) \sqrt{\frac{tT_i}{2r}} \tag{2)-(1a}$$

The economy of transaction balances increases with a decreasing number of monies (cf. Swoboda, 1968, pp. 39-41).

[4] In this "qualitative" sense money in comparison to barter could never be "neutral". Consequently, the concept of money neutrality is only conceivable in the quantity-theoretical sense: Money is neutral when an increase in the quantity of money leads to an equi-proportional increase in all money prices:

> "We had, so to speak, *qualitative* and *quantitative* theories of money. According to our qualitative theory, money was not neutral; it made a big difference. Pity the country that was still dependent upon barter, for it would have an inefficient economic system. But once this qualitative advantage had been realized by the adoption of the market structures using M, the *quantitative* level of M was of no particular significance (except for indicated transient states and uninteresting resource problems involved in gold mining or mint printing)." (Samuelson, 1968, p. 3.)

[5] ". . . a private institution which must issue money in competition with others, can only remain in business if it provides the people with a stable money which it can trust. The slightest suspicion that the issuer was abusing his position when issuing money would lead to a depreciation of its value and would at once drive him out of business. It would make him lose what might be an extremely profitable kind of business." (Hayek, 1984, p. 30.)

[6] "A reasonable explanation of why credit money did not replace commodity money before it did may not be because someone did not happen to think of the credit money idea, but rather may be because commodity money was, at the time, the cheapest way to produce confidence; i.e. a forced movement from commodity to fiduciary money would have implied a negative social saving." (Klein, 1974, p. 435.)

[7] Other (relatively recent) monetary experiences can be cited to underline the investment costs of new monies for the creation and maintenance of monetary confidence.

1. Currency reforms as a means for confidence renaissance in hyperinflating countries (as those of some European countries like Germany after World War I) could be considered as relatively easy to implement to the extent that the credibility of the existing currency had been completely eroded by the hyperinflation such that the newly created currency could not be worse than the former one. However, since the aim of a currency reform is to stop the inflation, this money has to be equipped with full monetary confidence. Consequently, one necessary (but not sufficient) condition for the creation of monetary confidence is the return to a commodity standard. Thus, the new currencies in the inter-war period were backed by gold or (other) international reserves.

   The *initial* investment cost for the establishment of a new and fully convertible currency is relatively low in terms of the necessary stock of gold or international reserves. Since, in a hyperinflationary economy, the *real* value of existing cash balances

[7] is at an extremely low level, their conversion amount into the new currency fully backed by (owned or borrowed) international reserves is also negligible. When, after the currency conversion, the currency reform (with its accompanying monetary and fiscal stabilization programme) disseminates increasing monetary confidence, the subsequent increase in the real demand for the new money can be satisfied by the monetization of additional international reserves coming from balance-of-payments surpluses (along the lines of the monetary approach of the balance of payments). (See Dornbusch, 1985, pp. 17-78.)
2. The creation of new national currencies by less developed countries after their access to political independence represents another example of the "commodity investment" cost of new monies. On the one hand, their quality had to be equivalent to that of the formerly circulating currency of the mother country. The newly created currencies had to be fully backed by reserves. On the other hand, the maintenance or intensification of monetary confidence necessitates the maintenance of fixed exchange rates, at least over some decades from the time when the new national monies were created.
3. When in the late 1970s the plans for the political independence of the province of Quebec from the rest of Canada were discussed, the most critical point was that of the monetary independence of Quebec. By recognizing the high resource costs for creating monetary confidence in a new money, the proposition by the "indépendantistes" (violently rejected by the other provinces) was that of the maintenance of the Canadian dollar for a politically independent Quebec (the plan "indêpendance-association").

[8] In the case of bankruptcy of a bank, the insurance company by reimbursing the depositors may also be forced to borrow from the central bank – in the case of insufficient reserves – and to serve its debt service afterwards out of future insurance premiums.

[9] A particular free-rider phenomenon from banks arises in the case where the lending banks have formed a syndication. When a loan to a debtor is of considerable size, the creditor side tends to be composed of a larger number of banks who, for reasons of risk diversification, join a syndicate. The formation of a syndicate is observable, in particular, in international lending operations where hundreds of banks constitute a syndicate. To the extent that the debtor becomes illiquid and that there is either suspicion of insolvency about the debtor or – as Sachs (1984) has shown – the necessity of rescheduling the debts at below-market rates in order to avoid a default by the borrower, each creditor has an incentive to free-ride in order to escape with its credit intact if the others renew the loan agreement. In particular, such a free-rider behaviour may be observable with respect to banks with small participation in the syndicate who then rely on the larger banks to forestall default.

[10] In this context the question could be asked why there had not been installed already a generalized deposit insurance system before the introduction of the compulsory one. The answer runs probably in terms of costs and benefits. To the extent that the banks relied on the already existing institution of the LLR function of the central bank, there was no additional motive for them to take part in a private insurance system since that would have been more costly for them.

[11] "Until recent years, true hyperinflation has occured only in countries undergoing revolution or severe civil unrest or that have been defeated in a major war, with the possible exception of John Law's experiment of doubling the French bank-note issue in the four-year period 1716 to 1720. However, currently, several countries seem on the verge of hyperinflation under relatively peaceful circumstances – Bolivia, Argentina and Israel, to mention only the most prominent. The misfortune of these countries promises to provide us with some evidence on a so far rarely observed phenomenon." (Friedman and Schwartz, 1985, p. 17.)

# Adjusting to Competitive Depression: The Case of the Reduction in Working Time

JEAN-PAUL FITOUSSI
and
DANIEL SZPIRO
Paris/Florence

Worsening unemployment in an inflationary context has weakened the credibility of growth as a policy for solving the employment problem. The non-cooperative character of international relations, which exacerbates the external 'constraint', leads governments, whatever their initial intentions, to adopt, more or less explicitly, restrictive policies. To increase unemployment becomes a substitute for protectionism, an instrument serving a higher-order objective: balancing external trade and defending the exchange rate.

On the global level, however, the crisis is of a Keynesian nature: the restrictive policies pursued in most Western countries have been combined notably with the fact that in the newly industrialized countries purchasing power corresponding to the recently installed production capacity has not been created. On the level of a developed country, it can be analysed in terms of insufficient competitivity (excessive wage costs) and/or of structural maladaptation of the productive apparatus. The fact that most developed countries are arguing along the same lines should have drawn attention to the inadequacy of that argument or at least encouraged more cooperative economic policies. The present strategy leads to a 'low pressure' equilibrium in the world economy, which implies that any attempt at recovery in a single country is condemned to failure. The existence of a reserve of unemployment seems inescapable. According to a recent memorandum from the Commission of the European Communities, stabilizing the rate of unemployment at around 11% of the working population (12 million unemployed) would need a growth rate above 3%. This seems impossible in the short term and unlikely in the medium term.

In such a context, with investment possibilities reduced by low growth prospect, the problem of employment becomes that of a voluntary or constrained reduction in the supply of labour. An arithmetical illusion

makes it look as if 8 men working 6 hours should give the same productive contribution as 6 men working 8 hours. Hence the idea of substituting one type of rationing for another. On the labour market, the existence of unemployment implies that the economic system spontaneously chooses one particular rationing scheme: those offering their labour are either totally satisfied or totally excluded from productive activity. Economic theory is practically silent on the reasons for this choice. Its rationality can be founded only on arguments which are generally exogenous to the models used. In the present state of economic theory, any type of rationing whatever might be possible, in particular a proportional reduction in working hours. There would then be no totally unemployed people, but all individuals would be partly unemployed if the supply of labour was in excess. It would therefore be possible to fix working hours in such a way that the whole available active population would be employed, i.e. in such a way that the disequilibrium of the labour market would be shared among all instead of borne by a few. Unemployment would then have an entirely different nature; it would no longer be an attribute of the individual, but a characteristic of the system as a whole.[1]

This solution implies the meeting of at least three conditions:

(i) that work is homogeneous and men interchangeable, otherwise working hours should be differentiated according to the respective availability of different categories of workers;

(ii) that production processes actually allow a 'substitution' between labour services and labour force;

(iii) that the wage fund is redistributed between the unemployed and those at work, i.e. that sharing out work is not simply a means of increasing the leisure of the active employed population. If the generous idea of sharing work serves as a pretext for redistributing the national income between the actively employed and employers, with the unemployed left out entirely, that had better be said. But then one should not in the same context deplore the lack of competitivity as the result of a growth in wage costs.

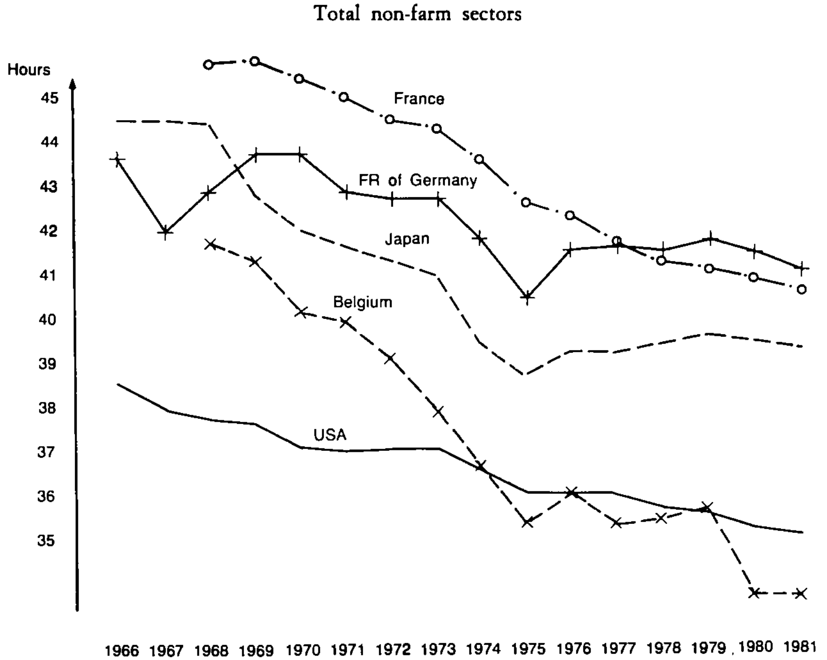
These conditions, particularly the last one, are imposed by the very premises of the reasoning. Long-term growth prospects are limited, and labour hence becomes a relatively abundant factor. "At the great feast of nature there would not be enough places." The Malthusian illusion of the excessive numbers of men raises its head again. But there are no lessons from history here, or at least they are ambiguous. The reduction in working time has characterized economic and social evolution since the industrial revolution. (Let us remember the times of Engels, where children's working hours had to be regulated to stop them exceeding 12 hours a day.) If this sharing out of work was possible yesterday, why should it not be tomorrow?

Since 1965, the trend to a progressive reduction in hours worked has

*Figure 1*

*Evolution of weekly working hours*

Total non-farm sectors




*Source*: ILO labour force statistics.

been general in all countries. A reduction in the legal working hours would, in this reasoning, only anticipate the trend and hasten the necessary adjustments in favour of employment.

But it is one thing to take note of spontaneous developments that take place against a background of economic and social progress. It is another to seek to constrain these developments and to justify this constraint by the absence of prospects for progress.[2] The reduction of working hours is in itself and end of our economic system, but it is doubtful that it constitutes a means towards a better distribution of scarcity in a crisis period. The conditions set out above are, of course, very unlikely to be met. The study of the consequences of the reduction of working time does not therefore proceed by simple arithmetic, but calls for a consideration of the whole set of interdependencies that characterize the functioning of the economic system. In practice, this means that econometric models should

be used for a simulation of the effects of the reduction in working time. Such an exercise, however, proves difficult and doubtful, for two fundamental reasons. The first is that the modelling of the production processes and their time dimension is still a largely unexplored area of economic analysis. The second comes from the fact that a reduction in working time represents a structural change – social processes are in general not reversible – and one does not know how it will affect the behavioural relations in the model. The reliability of the simulation results is therefore less the greater the proposed reduction in working hours.

## I - Econometric models

Econometric models cannot in general endogenously determine the direct effects of a reduction in working hours on the production process. The ambiguities in theorizing, combined with the lack of sufficient examples from the past – which were in any case in a very different economic context – explain this difficulty. A reduction in the legal hours of work thus remains outside the field of observation and experimentation of the models. The direct effects are therefore a priori the object of alternative hypotheses, which constitute different variants for study:

(i) What is the effect on productivity of changes in working hours?

(ii) Does the reduction in working hours entail a reduction in the duration of capital utilization and productive capacity?

A third a priori hypothesis relates to the conditions surrounding the reduction in working time, i.e. mainly to wages:

Is the reduction in working time accompanied by a sharing of income – i.e. a reduction in monthly income with unchanged hourly wages – or does it instead lead to an increase in wage costs? It is, of course, understood that wage income cannot undergo any reduction if the decrease in working hours is a substitute for an already planned or foreseeable wage increase, i.e. if it leads the workers to renounce other claims.

On the basis of a priori hypotheses concerning these effects and conditions, many simulations have been carried out in most European Community countries. Apart from the quantitative differences in result – the models used have different specifications, even if they derive from the same theoretical line – they alow a number of significant conclusions to be drawn.

(1) Employment always increases less than proportionately to the reduction in working hours. In other words, the total number of hours worked in the economy diminishes, whatever the hypothesis considered. In all simulations done for France using the DMS model, the one corresponding to a maximum effect on employment provides for a percentage

increase in employment less than half that in the reduction in hours worked. The gain in employment corresponding to a progressive shortening of the working week from 40 to 35 hours would on the most favourable conditions be 880 000 instead of some two million at the end of five years.[3]

(2) When account is taken of all the likely effects brought about by a reduction in working hours – including corrective policies applied to reduce the external disequilibrium – the models have brought out that the effect on employment is more favourable the closer one comes to the conditions set out in the introduction: maintenance of hours of utilization of capital, i.e. of productive capacity; and reduction in monthly wage income proportionate to the reduction in working hours. The models generally assume no long-term productivity gains. If however, despite these favourable conditions, employment does not rise in proportion to the reduction in working hours, the reason lies chiefly in two linked effects:

(i) The compensatory recruitment is not immediate, which, combined with the assumption of maintenance of capacity, implies a transitory increase in productivity, known under the name of productivity cycle.[4]

(ii) It follows that the wage fund, demand and production costs undergo a relative reduction, as does the growth rate of the gross domestic product. The share-out of work and income does not therefore take place identically, since production shows a slight decline (varying according to model, but always present).

(iii) On these hypotheses, shortening working hours becomes a powerful means of reducing both the disequilibrium in welfare budgets (particularly the social security deficit) and the external deficit.

(3) The models differ more in quantitative evaluation when the hypothesis of wage compensation is considered. This implies, immediately following the cut in working hours, an increase in hourly wage rates, and with productive capacity unchanged, an increase in wage costs which is higher the less the transitory growth in productivity. The increase in wage costs is in fact exactly equal to the costs per unit poduced of the extra recruitment carried out by the firms. The wage compensation thus brings about two contradictory effects, which explain the divergences among the models. On the one hand, the increase in the wage fund leads to an increase in effective demand. On the other, the growth in wage costs damages the competitivity of firms if it leads to inflation, and reduces their profitability if it does not. Depending on whether the models favour the first or the second of these effects, they will deduce overeffectiveness of the measure for employment (as in the Metric model) or an attenuation in its positive effects (as in DMS). The following table gives a picture of the divergences between models from different countries.

Independently of these divergences, which relate to the different specifications of certain relationships, particularly the investment function, the

*Table 1*

Effects on employment of a reduction in working hours, on various hypotheses of wage compensation

| *Country* Hypothesis | No wage compensation | Wage compensation |
|---|---|---|
| *France* | | |
| Metric | + | + + |
| DMS | + | + − |
| Propage | + | + + |
| *FR of Germany* | | |
| Institute for planning and decision-making systems | + | + + |
| Universities of Tübingen and Freiburg | − | − + |
| *Belgium* | | |
| Planning Office | + | + − |
| *Denmark* | | |
| Finance Ministry (ADAM) | + | + − |
| Economic Council | + | − |
| *Netherlands* | | |
| Economic and Social Council (Vintaf II) | + | − |
| University of Amsterdam | + | + |
| *United Kingdom* | | |
| Treasury model | + | + − |
| *European Community* | | |
| Comet III | + | + |

A plus (+) in the first column indicates a positive effect on employment; a minus (−) a negative effect. A + + in the second column means an accentuation of the positive effect; a + − a reduction in this effect with the overall effect remaining favourable. Where only one sign appears in the second column, it indicates either that the overall effect of the reduction in working hours is negative on the assumption of wage compensation (sign −); or that this effect remains positive but is more or less the same whether there is wage compensation or not.

simulations generally do not take account of feedback effects of the evoltion of the economy on economic policy.[5]

The hypothesis of wage compensation can be analysed in relation to the first column of the table, as a measure of 'recovery' of economic activity through consumption. Its success or failure (always in relation to the

opposite hypothesis) will depend on whether the external equilibrium has contraining force or not, i.e. whether a worsening in the trade deficit will bring about a corrective policy or not. If that were the case, this policy would cancel the positive effects expected from the measure, since it would use as an instrument the reduction of effective demand. The fact that the 'external constraint' bears differently on the countries considered probably explains the divergences in their conclusions. One need not therefore be astonished that wage compensation has a positive effect in the Federal Republic of Germany.

When on the other hand the external balance has constraining force, i.e. when the government's financial and monetary policy is endogenized, a 'trade-off' would exist between the change in real wages and employment.[6] This trade-off implies that any increase in real wages takes place to the detriment of employment.

(4) Since it is an object of negotiation, the question of wage compensation can theoretically be dealt with. That of the effect of reduction of working hours on the duration of utilization of capital is of quite a different nature, and implies a knowledge of the determinants of production which is still only fragmentary.

The reduction in the duration of utilization of capital has clearly negative effects on employment in practically all simulations. The numbers of men employed always grow less than on the assumption that productive capacity is unaffected, and even sometimes fall by comparison with the reference situation where legal working hours are not changed. If in fact production fell, firms' employment needs would not be changed by the reduction in working hours, but unit capital costs would increase whether or not there were wage compensation. The effects on the balance of trade would obviously be negative, because of both a reduction in exports and an increase in imports.

By way of illustration, the following table describes what would, according to DMS,[7] be the effects of a reduction in working hours combined with a reduction in the duration of capital utilization in the sector exposed to foreign competition. The hypothesis considered is that of a reduction from 40 to 35 hours per week, in steps of one hour per year.

The decrease in the duration of capital utilization seems to cancel out almost entirely the positive effects on employment of a shortening in working hours, even on the assumption of no wage compensation.

We have deliberately confined this brief overview of the findings of simulations done using econometric models to their qualitative aspects rather than the quantitative ones. On the one hand, in fact, the extent of reduction in working hours contemplated in the various countries is not the same. On the other, the models used differ in their specifications, so that the reference situation – the forecast to which the consequences of the economic policy measure are compared – is not necessarily identical

*Table 2*

Effects of a shortening of working hours with reduction in the duration of utilization of capital (exposed sector) (*average annual changes over 5 years*)

| Economic indicator | Unit | Reduction of 1 hour per year in working hours | Reduction of 9/10 hours per year in hours of use of equipment | Combined effect |
|---|---|---|---|---|
| GDP in real terms | Average % per | −0.08 | −0.75 | −0.83 |
| Household consumption | year for the | −0.36 | −0.27 | −0.63 |
| Gross investment (firms) | period 1981-85 | +0.24 | −0.78 | −0.54 |
| Capacity (+) or need (−) for financing | | | | |
| — of administrations | Million FF | +9 793 | −12 155 | −2 362 |
| — of the Nation | (1980) per year | +6 714 | −2 282 | +4 432 |
| Growth in consumer prices | Average % per year | −0.24 | +0.56 | +0.32 |
| Growth in purchasing power | | | | |
| — Annual net gain for wage workers | Average % per year | −0.75 | −0.16 | −0.91 |
| — Disposable household income | | −0.40 | −0.27 | −0.67 |
| Change in total employment | Thousands, per year | +74.3 | −70 | +4.3 |

*Source*: Preparatory work on VIIIth Plan, "Rapport du Comité Emploi-Revenus, La Documentation française", 1980, p. 157.

from one simulation to another. A quantitative comparison would imply a prior distinction of what in the differences in evaluation of the consequences of shortening working hours is due to a differing representation of economic relations.

It would, however, seem that the findings of simulations as to the direction of the effects on employment of a reduction in working hours are comparable. Everything depends on three parameters: the proportion in which the reduction in working hours leads to lasting gains in productivity; the proportion in which it leads to a fall in productive capacity; the proportion in which its cost is supported by firms. If these three proportions are zero, the effect on employment is maximum. The models generally assume that the first is zero, with the productivity gains envisaged being only transitory. A consideration of positive values for the second and/or third proportion generally leads to an attenuation of the positive effects on employment. The least favourable configuration is obtained when both are

equal to 1. But it is precisely the fact that these parameters are unknown that constitutes the problem. Attempts at quantification have therefore not reduced our uncertainty as to the consequences of a policy of shortening working hours.

## II - Working hours, duration of the production process and production capacity

The hypothesis of maintenance of productive capacity used in the simulations in fact assumes that the problem is solved, since it implies a substitution between hours worked and men employed. It therefore has the immediate consequence of a growth in the demand for labour in terms of men.[8] In other words, the shortening of working hours is credited from the outset with a favourable effect. Although there is no lack of ingenious arguments to justify the assumption, there is equally none of ones to contradict it. But their generally ad hoc character damages the credibility of the argument. This is why it is useful to turn towards theoretical considerations.

When one considers a general representation of production processes the idea of substitution between working hours and men at work, i.e. between labour services and labour force appears superficial. Such a representation implies a distinction between the 'agents' of the productive processes (capital, number of workers) and their services (working hours, duration of capital utilization). Production per unit of time is then a function of the 'agents' – so many machines combined with so many men produce so and so much product per hour – and total production in a given period is proportional to the duration of the production process. This cannot be confused with working hours unless there is only one shift. This theoretical outline allows some interesting conclusions to be drawn, which we shall only mention in the context of this article:[9]

(i) A reduction in working hours does not lead to the formation of an extra shift to keep up the production level unless it reaches a certain proportion defined by very strict conditions. If these conditions are met, the duration of capital utilization increases, with unchanged production. The transition to an extra shift is accompanied by a reduction in the number of men per shift, but total employment increases in a proportion equal on a first approximation to that of the reduction in working hours.

(ii) If the reduction in hours does not have the extent necessary for this critical proportion, production falls. In any case *at unchanged wage rates, optimal employment is independent of working hours.*[10] In order for productive capacity to be maintained, the fall in the duration of capital utilization would have to be compensated for by an increase in employment. But this

is only possible if the wage rate goes down. Income per worker would then undergo a reduction more than proportionate to that in working hours. The increase in the number of men employed when the duration of capital utilization is reduced implies in effect a decrease in labour productivity.[11]

One must of course step back somewhat from these theoretical results, especially because they proceed from statics and assume that capital is constant. But they allow a better understanding of the productive process because they take account of the relations between production time, working hours and capacity. They lead one to think that the most probable result of a policy of a progressive but slight reduction in working hours would be a fall in production capacity.

However, even if the reduction in working hours is below the critical proportion, this fall need not be proportional to that in working hours except in extreme circumstances which imply the meeting of at least three conditions:[12]

(i) There is no excess capacity ("the margins of production capacity with recruitment" are at their optimum level).

(ii) The reduction in working hours brings about no increase in productivity.

(iii) No investment is made to compensate for the fall in productive capacity.[12]

If one of these conditions is not met, the fall in productive capacity would be below that in working hours to an extent that depends precisely on the available margins of capacity, the gains in productivity or the investments. For instance, in the simulation done by the Danish Finance Ministry, the question of loss of capacity does not play an important part, since on the hypothesis that it takes place, it encourages sufficient investment to compensate for its negative effect on employment. This case is obviously extreme, but it draws attention to a causality that the models generally ignore, between reduction in working hours and investment.[13]

The proportion in which a relatively slight reduction in working hours ought to bring about a reduction in productive capacity should therefore be positive, though less than 1. Jacques Drèze has estimated, for the Belgian economy, that on the assumption that the cost of the cut in working hours is $\frac{2}{3}$ supported by the workers, this proportion ought to be les than $\frac{1}{4}$ in order for the effects on employment of the proposed policy to be favourable.[14]

## III - Working hours and productivity

The question of the 'dilemma' of productivity has been widely discussed in all the works dealing with the effects of shortening working hours. There

is a dilemma because the maintenance of productive capacity calls for an increase in labour productivity, but if this does take place the effects on employment are thereby attenuated. We have seen that the adjustment of actual to desired employment generates a productivity cycle, but that the resulting growth in productivity is in an inverse relation to the speed of the adjustment. In other words, if the firm proceeds instantly to compensatory recruitment, labour productivity would not change. In any case, this process is only transitory, and does not affect labour productivity in the long term. In most models, labour productivity is therefore not affected by a reduction in working hours.

However, there is a convergent series of studies concluding instead that there are lasting effects on productivity of a reduction in working hours,[15] and a survey done in 1978 by the journal *L'usine nouvelle* indicates that firms would initially if not primarily react to such a measure by trying in all ways to increase productivity. Other surveys have confirmed this finding.[16]

It is our intention here, without prejudging the direction of the causal links between changes in working hours and changes in productivity, to bring some theoretical considerations to bear on the relation between these two variables and to present the results of an empirical exploratory study.

Recent developments in the economic theory of labour conflicts have shown that there is a trade-off between the union claims regarding working conditions, primarily working hours, and those regarding wages. Increases in productivity allow both these claims to be partially satisfied.

In the neoclassical approach to the supply of labour, the expected growth in productivity and therefore in real wages brings about an intertemporal effect of substitution between work and leisure which may be reflected in a reduction in the individual supply of labour.

More specifically, since the end of the 1950s, the increase in hourly productivity of labour has in fact been accompanied by a fall in the annual working hours, compensated by a smaller growth in the purchasing power of the average wages. In this evolution, the increase in productivity therefore seems to be a factor determining the reduction of working hours.

The Marxist concepts of surplus value – Marx was one of the rare authors to deal with the problem of working hours – allow emphasis to be made on the existence of a trade-off between working hours and productivity, but this time from the point of view of firms.

The reduction in absolute surplus value (which depends on the length of the working day) leads firms to seek ways to increase relative surplus value, which varies in direct proportion to productivity. This means is, of course, the accumulation of capital, i.e. the investment. It is therefore the prospect of productivity growth that leads workers to demand and firms to accept a reduction in working hours.

Expectations of lasting productivity gains therefore seem to govern the spontaneous evolution of working hours.[17]

We have attempted a rough empirical test of this relationship. Medium-term expectations of productivity developments are not observational data, and we have formulated a hypothesis regarding their formation.

Productivity has a 'technological' component depending on the productive combination of the factors (equipment, labour skills) and the organization of production. We assume that the firm is prudent in forecasting future technology, because of a twofold uncertainty: firstly regarding technical progress, and secondly regarding its own financial situation, which may slow down implementation of productivity investments. Firms will expect productivity gains to continue on the basis of existing technical standards.

The second component of the expectations is of an economic nature. For a given technology, lags in adjusting the production factors to their desired level determine short-term fluctuations in productivity. It is our hypothesis that firms form their expectations of productivity on the basis of a normal rate of utilization of productive capacity. It is therefore potential productivity rather than actual productivity that constitutes our reference.

The relationship between working hours and expected productivity has been compared with statistical data from three countries (France, Italy, United Kingdom). The econometric results and the method of estimation used are described in the annex. The partial nature of the relationship and the uncertainties related to the international statistical comparisons prevent us from proceeding to draw quantitative conclusions on the basis of these results.[18] But our aim was not to make quantitative forecasts on the basis of these relationships, but to look for a confirmation of a hypothesis, mainly that there is a structural relation between working hours and productivity in the medium term; and such relation does seem to exist in the three countries.

Whatever the causal links between these two variables, and our hypothesis favours that going from economic growth to working hours, a reduction in legal working hours will probably lead to lasting productivity gains. But this conclusion is not so negative as it seems regarding the effects of such a measure on employment. One cannot in fact lose on both grounds; an increase in productivity ought to partly compensate for the negative effects of the reduction in working hours on productive capacity.

In any case, neither one of these effects is independent of the evolution in labour cost since the latter in part determines both the developments in productivity, throuh its influence on investment, and that of profitable capacity, at least in the short term.

## IV - The question of wage compensation

The question of the increase in wage rates is clearly not independent of the diagnosis one makes on the present economic recession. If there exist excess capacities, if the problem is that of the failure of effective demand, then a shortening in working hours will have the more effect on employment if it is accompanied by an increase in wage rate, i.e. by total wage compensation.

If instead the diagnosis is that of a shortage of capacity, of a lack of profitability, then wage compensation will lead at best to inflation and devaluation, at worst to a worsening of unemployment.

But a policy of shortening working hours is better founded on the second diagnosis, since it assumes that employment (in terms of hours worked) has reached a level that is hard to exceed, and that therefore unemployment must be shared differently among workers. As we have stressed, this policy amounts to substituting one rationing scheme by another; to acting on the effects rather than on the causes. If this diagnosis is well founded, other policies are possible and perhaps preferable,[19] since organizing the rationing of labour may have a number of consequences that must be made explicit.

The reduction in working hours and proportionately of income is unacceptable to the worker unless the marginal 'disutility' of labour compensates exactly for the utility of the fraction of income that he is giving up. Were this not the case, were the first less than the second, the worker would seek means of loosening the constraints facing him. The risk of seeing a dual labour market emerging is therefore associated with such a policy. But this risk does nothing more than replace another one: the existence of a sizeable number of completely unemployed people is likewise a powerful motor of the underground economy. A second consequence may be an increase in demand for part-time employment; but the unions are opposed in almost all European countries to an expansion in part-time work, since it would accentuate the discrimination between a primary and a secondary labour market.[20]

The union strategy is precisely to avoid such effects by calling always and everywhere for total wage compensation. This is part of a rigorous logic aimed at preventing expansion of an ill-controlled area of the labour market, and on the other hand at combating a policy of passive adaptation of the supply of labour to employment. "It must be recognized that the reduction in working hours is not genuinely a collective demand raised by the majority of workers."[21]

An approach to the condition set out previously implies both partial wage compensation and a contraction of wage scales. The impact of these measures on firms will be differentiated according to the employment

structure necessary for their functioning. Another possibility is for the State budget to take over all or part of the wage compensation.

Shortening working hours therefore generally implies compensating, at least partly but also in a differentiated manner, for its effects on workers' incomes. Recent French experience shows this well, difficult as it may be to know what the wage increases might have been if the working week had not been cut by an hour. In a dynamic perspective, it is the difference betwen this hypothetical situation and the actual change in remuneration that determines the extent of the compensation.

If it is true that the present situation can be analysed as a situation of classical unemployment ('too' high real wages), then a better alternative would exist to a shortening of working hours. A reduction in wage rates would have pösitive effect on both employment and production. The problem is knowing what is the elasticity of employment in relation to real wages. If this elasticity is too small, such a policy would be completely impracticable, since the wage cut needed to restore full employment would be too great.[22]

But is this diagnosis well founded? Recent work by the OECD and the EEC seem to prove this. They use a concept created ad hoc, 'the real wage gap', regarded as measuring the difference between the real wage and its equilibrium level. This gap is supposed to be notably positive for Belgium, France, and Italy... but also for Japan. Serious theoretical and empirical doubts may therefore be cashed on the significance of this gap as a measure.[23] Another diagnosis might be formulated, leading to a different policy.[24]

Our rapid survey of the problems raised by cutting working hours *as an employment policy* is probably too negative. This is due more to the philosophy underlying this type of measure than to the 'scientific' knowledge of its consequences. Simulations done using econometric models ought as we have seen to be analysed with circumspection. The uncertainty in their quantitative evaluations is because they are based on prior calculation of the direct effects of shortening working hours; but these effects are not very well known. Some of its indirect effects cannot, moreover, be taken into account. Would a considerable reduction in working hours bring with it a change in consumption patterns?

This uncertainty is combined with those regarding the diagnosis one makes on the situation of the economy, and regarding the probable development of its international environment.

Shortening working hours is a policy consisting in adapting to prospects felt to be bad for the future, rather than trying to modify them. It also pertains to a new system of international relations, 'competition through depression'. It is our convinction that the context in which it is implemented – adjustment of the supply of labour rather than of the employment – leads wage earners to percive it as a constraint. Its ambition is not

to absorb unemployment but to distribute it differently. Things would be different if it were taken in the context of a positive sum game. But precisely its chances of success are subordinated to the hypothesis of no increase in productivity.

Alternatively, one might regard our societies as having become sufficiently rich to allow for less work. Why then impose a redistribution of income among workers alone? Why not think of a bolder redistribution not only of income but also of wealth among all social categories?

## Bibliography

G. Basevi, O. Blanchard, W. Buiter, R. Dornbush and R. Layard, "Macroeconomic prospects and policies for the European Community", *Economic Papers* No 12, EEC, April 1983.

Commission of the European Communities, "Memorandum on the reduction and reorganization of working hours", December 1982.

Y. Barou and J. Rigaudiat, *Les 35 heures et l'emploi*, La documentation française, 1983.

Y. Barou, J. Rigaudiat and A. Doyelle, "Réduction de la durée du travail et comportement des entreprises. Synthèse des études monographiques disponibles", *Travail et emploi*, July-September 1982.

Y. Barou, F. Perronet et F. Rocherieux, "Réduire la durée du travail. Analyse macroéconomique et enjeux sociaux", *Travail et emploi*, July-September 1982.

E. Denison, *Why growth rates differ*, The Brookings Institution, 1967.

J.H. Drèze, "Salariés, emploi et durée du travail", *Recherches économiques de Louvain*, March 1979.

J.H. Drèze, "Réduction progressive des heures et partage du travail", Intervention at the Fourth Congress of Belgian French-speaking economists, 1980.

J.H. Drèze and F. Modigliani, "The trade-off between real wages and employment in an open economy, *European Economic Review*, 1981, No 1, Vol. 40.

J.P. Fitoussi, "Politique monétaire ou politique économique?", *Observations et diagnostics économiques*, No 1, June 1982.

J.P. Fitoussi and D.P. Nuti, "Indicizzare i salari alla disoccupazione?", *Politica ed economia*, November 1982.

J.P. Fitoussi and N. Georgescu-Roegen, "Structure and involuntary unemployment", in E. Malinvaud and J.P. Fitoussi (eds.), *Unemployment in Western countries*, MacMillan, 1980.

INSEE, "Maximiser l'emploi. La réponse de Metric", November 1980.

E. Malinvaud, "Une explication de l'évolution de la productivité horaire du travail", *Économie et statistique*, September 1973.

G. Tahar, "Durée du travail, investissement et emploi: étude monographique", Toulouse CEJEE, University for Social Sciences, 1981.

G. Tahar, "Combinaison productive et durée du travail", *Travail et emploi*, July-September 1982.

## Notes

[1] Other methods of rationing are obviously possible and used: lowering the retirement age, extending compulsory schooling, lengthening annual holidays, etc. Though they do not have the same consequences as those of cutting working hours, they proceed on the same

philosophy: adjusting the supply of labour to employment instead of employment to the supply of labour.

[2] The effects on employment of the 1936 introduction of the 40-hour week are controversial (cf. A. Sauvy). This was a great and necessary social conquest; but while it may be legitimate not to separate the economic from the social, it is not legitimate to seek at any cost to justify a piece of social progress that is good in itself by doubtful economic benefits.

[3] These simulations were done as part of the preparatory work for the plan. Cf. Y. Barou and J. Rigaudiat (1983; Chapter 2).

[4] The models assume that the recruitment behaviour of firms is similar whether they have to cope with the needs of an increase in demand or those of a reduction in working hours.

[5] There are of course exceptions. Cf. INSEE (80a).

[6] Cf. J. Drèze and F. Modigliani (1981).

[7] A shortening in working hours implies, on the DMS model, a coresponding reduction in productive capacity, while in Metric it is entirely compensated for by a growth in labour productivity (cf. G. Oudiz, E. Raoul and H. Sterdyniak, 1979).

[8] In the DMS model, the growth in the demand for labour is more than proportionate to the reduction in hours worked, since the maintenance of productive capacity calls for the use of older equipment.

[9] For a formal demonstration, refer to the article by J.P. Fitoussi and N. Georgescu-Roegen (1980).

[10] It is only because authors habitually confuse labour force and labour services in representing the production function that this result is unknown. It recalls Karl Marx's objection to Nassau Senior's assertion that profit was earned during the last hour of the working day. For Karl Marx, it was instead proportional to the total number of hours worked.

[11] As in vintage models such as DMS.

[12] On this point see, in particular, Y. Barou, F. Perronet and F. Rocherieux (1982) and Y. Barou and J. Rigaudiat (1983).

[13] The inverse connection is clearly better known and more studied (cf. G. Tahar, 1981, 1982).

[14] Cf. J. Drèze (1980).

[15] Cf., in particular, E. Denison (1967) and E. Malinvaud (1973).

[16] Cf. Y. Barou, J. Rigaudiat and A. Doyelle (1982).

[17] This relationship may, of course, be given another interpretation. Shortening working hours may be the means or employers to secure acceptance for a further increase of productivity (cf. Y. Barou, F. Peronnet and F. Rocherieux, 1982).

[18] Our method of estimation has moreover not allowed elimination of the autocorrelation of remainders in the relationship estimated for the French data.

[19] Cf. J.P. Fitoussi and D.M. Nuti (1982).

[20] Cf. Commission of the European Communities (1982).

[21] Journal CFDT *Aujourd'hui*, January 1982.

[22] Cf. J. Drèze (1979).

[23] Cf. G. Basevi, O. Blanchard, W. Buiter, R. Dornbush and R. Layard (1983).

[24] Cf. J.P. Fitoussi (1982).

# Inflation and Interest: The Fisher Theorem Revisited

Wolfgang Gebauer
Frankfurt/Florence

> *"...the rate of interest is always relative to the standard in which it is expressed"*
> *(Irving Fisher, 1930)*

## Introduction

Towards the end of the nineteenth century, Irving Fisher undertook a theoretical and empirical examination of the general relationship between interest rates, expressed in different standards, and expectations of changes in the value of these standards. He formulated specifically the hypothesis of an immediate, direct and full adjustment of nominal interest rates on financial assets to expected changes in the purchasing power of money. This hypothesis implied constant real interest rates that were not influenced by inflationary expectations.

In the period since the end of the Second World War, the Fisher theorem has been "rediscovered". The renewed interest has been due in particular to the modern debate on monetarism and the similar "rediscovery" of the quantity theory of money, and to the phenomenon of persistent inflation in Western industrialized countries. The relationship between *Appreciation and interest* (the title of Fisher's pioneering work in 1896) has become a central element of modern monetary theory in the form of long-term "neutrality" of money and inflation in relation to real variables (such as output and employment); this "neutrality" implies constant real interest rates.

The Fisher theorem has thus been incorporated into major areas of modern economic theory as applied to the inflation phenomenon. It is used in analysing individual financial markets in a national and international context; it is part of microeconomic, intertemporal price theory and part of the macroeconomic theory of investment and fluctuations in the business cycle; and it quite generally underlies the practical and economic

policy calculation of real interest for individual markets or economies, using real variables adjusted for price changes.

Yet the way in which the Fisher theorem (or Fisher equation) has hitherto been received reveals some confusion and analytical shortcomings. The restrictions applying to the original, complete Fisher theorem are given scant attention in the literature; the expectation formation hypothesis ascribed to Fisher is a scarcely tenable assumption; the two-way link between interest and inflation, which Fisher emphasized from the outset, is consistently ignored. This "list of sins" could be extended, as this paper will show.

Any critical review of past theoretical and practical interpretations of the Fisher theorem must start from a historical-analytical examination of Fisher's original writings. Chapter 1 therefore describes the original theorem and gives an outline of Fisher's concept of interest (Annex to Chapter 1). Chapter 2 traces Fisher's three-stage, empirical procedure and summarizes his findings. A crucial point here is that leaving aside basic methodological and empirical questions, Fisher's factual evidence quite clearly falsifies his original theorem. The conclusions which Fisher drew from his factual evidence are examined in Chapter 3; they relate to Fisher's explanations of the Gibson paradox and of the long time lags he had detected, to his statements on the reversed causality between interest and inflation and interest-rate policy.

## 1. Fisher's original theorem

### 1.1. Starting point and focus

The starting point of Fisher's investigations into the subject of "interest and inflation" was the bimetallism controversy, i.e. the monetary policy argument in the last quarter of the nineteenth century over how international reserve assets should be defined.[1] Because of the appreciation of gold, which meant a fall in prices, the bimetallists claimed that debtors had been robbed, since, on the gold standard, the rising purchasing power of gold was tantamount to discrimination against debtors of contractually fixed amounts of money (gold) payable in the future. Fisher attempted to settle the controversy which arose with the monometallists (the advocates of a pure gold standard as opposed to a gold and silver standard) by pointing out in his monograph *Appreciation and interest* (1896) that the essential element in a loan contract was not (as the bimetallists asserted) the scaling of the (principal) sum owed (for example in the standards gold or silver), but the interest rate agreed between debtor and creditor. For, so Fisher argued, an expected appreciation of gold would change the interest rate reckoned in that standard such that only the nominal amounts of

money, but not the quantities of goods represented by the amounts of money would be affected,[2] given certain ideal assumptions (see next section).

The idea of (market) interest rates adjusting to expected changes in the value of the relevant standard (unit of account) was developed by Fisher from suggestions put forward earlier by J.S. Mill, de Haas and Clark.[3] This basic idea is also evident in Fisher's agio concept of interest, whereby the interest rate is seen as the percentage premium on present goods when exchanged for future goods of the same kind and number.[4] However, the nature of the goods (like the time-span between present and future) influences this perentage premium. Consequently, interest rates in different standards (for different types of goods) differ basically from one another.

On the basis of this idea, Fisher saw his first task in formulating in exact theoretical terms the relationship between those inherently different interest rates and expected changes in the value of the various standards. He concentrated on the relationship between two particularly interesting "typical" interest rates, namely a nominal interest rate in the monetary standard (unit of account "money") and a "real" interest rate in commodity standard (unit of account "goods", based on a specified basket of commodities). It is above all this relationship between nominal and real interest rates which was then investigated empirically by Fisher and applied to economic policy issues.

## 1.2. Assumptions and definitions

Fisher's theory on the relationship between interest and inflation is based on a number of highly simplified assumptions about expectations, market behaviour and the way in which financial markets operate. These assumptions sketch an idealized world in which there are no risks or uncertainties.

*Assumption 1:* On all financial markets, expectations are formed in respect of the future relationship between the value of money and the value of goods. A change in the future value relationship between money and goods is uniformly expected by all market participants: they anticipate a specific rate of price change $p^*$ (e.g. a specific inflation rate).

*Assumption 2:* The future rate of price change $p^*$ (the asterisk denotes an expectation variable) is anticipated with "perfect foresight": market participants are never mistaken in their price expectations. Since they all possess perfect foresight, they also form the same inflationary expectations.

*Assumption 3:* Financial markets are characterized by perfect transparency and the absence of any risk considerations or transaction costs. There is, for example, certainty as to payment and repayment of loans and

monetary assets at predetermined points in time. All market interest rates are therefore "pure" interest rates in the sense that they contain neither a risk premium nor a transaction cost component.

Assumptions 1 and 2 show that Fisher's theory relates to expected changes in the price level and hence to expected real interest. The way in which expectations are formed and the foresight associated with these expectations are, therefore, a main focus of theoretical and empirical attention. If, like Fisher, one assumes perfect foresight, expectations can never be wrong: there are no unexpected rates of price change, because the actual rate of price change is always predicted correctly. Consequently,

$$p^* = p.$$

Obviously, the assumption of perfect foresight, implying $p^* = p$, means considerable abstraction from economic reality, one main feature of which is, after all, uncertain price expectations. On the lines of Fisher's work on the quantity theory of money,[5] perfect foresight corresponds to the situation of long-term market equilibrium; this contrasts with disequilibrium transition periods where there are imperfect or uncertain expectations (see Chapter 3 below).

Fisher's theory on "interest and inflation" embraces positive and negative expected rates of change in the price level, i.e. inflationary and deflationary expectations. The following definitions and terms are tailored to the existence of inflationary expectations in the sense of an expected rise in the value of goods relative to the value of money. The units of goods in a particular shopping basket (e.g. the basket for all private households) are measured using the price index P. P indicates the value of the shopping basket in units of money (the "price level"). The expected price index (for example one year hence) is written as $P_1^*$, and the present price index as $P_0$. The expected relative (percentage) change in the price index in the coming year, i.e. expected inflation rate $p^*$, is accordingly defined as

$$p^* = (P_1^*/P_0) - 1\text{[6]}.$$

The symbol $p^*$ thus denotes the expected change in value, measured in units of money, of the goods in a particular shopping basket. Where there are inflationary expectations ($p^* > 0$), the value of the goods rises relative to the value of money, and vice versa where there are deflationary expectations ($p^* < 0$).

The situation $p^* > 0$ can also be expressed using the "purchasing power of money" concept, which is inverse to the price index or price level: if there are inflationary expectations, it is anticipated that money will lose value relative to goods, so that a decline in the purchasing power of money is expected. This loss in the value of money, or expected rate of

change in the purchasing power of money $\pi^*$, is measured in units of goods.[7]

Fisher's theory could therefore be formulated both in terms of the expected relative change in the value of goods in units of money $p^*$ and in terms of the expected relative change in the value of money in units of goods $\pi^*$. The first formulation is opted for below, the rate of change of the price index in units of money being easier to capture statistically. The expected inflation rate $p^*$ is written in decimal notation (e.g. $p^* = 0.04$ to denote an expected inflation rate of 4%).

The actual nominal interest rate, i.e. the interest rate measured in units of money, which prevails or is agreed today and is to remain unchanged (for one year) is referred to as i; it is also, synonymously, called the market interest rate or money interest rate. Nominal interest rates are agreed between creditors and debtors on financial markets, particularly on the money market, the credit market and the securities market.[8] The agreements relate to the future contract period, for example to the agreed duration of a loan contract. The nominal interest rate as quoted on the financial markets is consequently an ex ante interest rate. Such an ex ante interest rate is, as it were, forward-looking: it has a time horizon set in the future. Nevertheless, it is not denoted below as an expectation variable (by an *), since it represents a contractually agreed variable fixed in the present. The expected nominal interest rate $i^*$, by contrast, is a variable whose contractual determination will take place only in the future.[9] "Expectation" is used here not as a contractual, but as a psychological notion, defined as "... attitudes, dispositions, or states of mind which determine our behaviour, or at least accompany it".[10]

Accordingly, the (ex ante) real interest rate is an expected magnitude, i.e. a variable that can be established only through future expected rates of change in the price level. This ex ante expected real interest rate may be expressed as $r^*$. For the purposes of Fisher's theorem, $r^*$ must be seen as the "true" or "real" interest rate, i.e. interest rate adjusted for price changes (measured in terms of the units of goods in a particular shopping basket) which is at present expected with perfect foresight (but not expressly agreed) for the coming year. Because of the congruence in time between expected inflation and the (implied) ex ante time horizon of the interest rate, the terms "ex ante" and "expected" can be used synonymously here. Given perfect foresight, the expected interest rate $r^*$ is, because of the implication $p^* = p$, always the same as the real interest rate r actually observed; i.e.

$$r^* = r.$$

Accordingly, the distinction, in terminology and symbols, between expected and observed inflation rates and real interest rates would seem to be superfluous. However, it is emphasized here from the outset, since it is

of prime importance later in the interpretation of Fisher's factual evidence and in the reformulation of his original theorem.

### 1.3. Derivation

In deriving the theorem, let us assume, according to Fisher,[11] a time horizon of one year with given nominal interest rates and perfect inflationary expectations. Now let us suppose that a certain sum of money D (currency units) and the quantity of goods W (units of goods) are lent today for one year. In one year's time, the repayment of $D(1+i)$ dollars will be due on the sum of money D lent today; at the same time, the repayment of $W(1+r^*)$ units of goods will be due on the quantity of goods lent today. The symbols i and $r^*$ stand for the nominal and real interest rates defined above. When the loan contract is concluded, the sum of money D is worth just as much as (equivalent to) the quantity of goods W, i.e. "today"

$$W \triangleq D.$$

To derive the theorem, we have to consider explicity an interest rate aspect and an inflation aspect of the one-period loan.

*Interest rate aspect*

To maintain equivalence between the sum of money D and the quantity of goods W after one period, the relative increase in quantities due to interest rate payments must be the same, i.e.

$$D(1+i) \triangleq W(1+r^*). \tag{0.1}$$

*Inflation aspects*

In the course of the year, the quantity of goods W is certain to rise in value relative to money, i.e. by the inflation rate $p^*$ of goods W in terms of money. After one year, therefore, a sum of money increased by the inflation rate, i.e. $D(1+p^*)$, is necessary in order to be able to buy the same quantity of goods as initially, namely

$$W \triangleq D(1+p^*) \tag{0.2}$$

or extended, for better comparison with (0.1), by $(1+r)$, we obtain

$$W(1+r^*) \triangleq D(1+p^*)(1+r^*). \tag{0.3}$$

The left-hand side represents the quantity of goods and the right-hand side the sum of money which, with a given inflationary expectation, is necessary to redeem the debt after precisely one year. This sum of money was already identified above as $D\ (1+i)$.

*Corollary*

Hence, if we look for the relationship between i and $r^*$ which assures equivalence between money D and goods W, we can directly derive, by comparing (0.1) and (0.3), the *equation*.

$$D\ (1+p^*)\ (1+r^*) = D\ (1+i).$$

After reduction by D, Fisher's original and complete theorem is obtained:

(1.1) $$\boxed{(1+i) = (1+r^*)\ (1+p^*) \quad p^* = p.}$$

Transforming (1.1) gives directly the original, complete nominal interest theorem (1.2) and the original, complete real interest theorem (1.3):

(1.2) $$\boxed{i = r^* + p^* + r^*p^* \quad p^* = p}$$ (nominal interest theorem)

(1.3) $$\boxed{r^* = (i - p^*)/(1+p^*) \quad p^* = p}$$ (real interest theorem)

In accordance with (1.2), the nominal interest rate i is determined by the ex ante expected real interest rate $r^*$, by the inflation rate $p^*$ expected with perfect foresight and by the product of these two variables $r^*p^*$, which is to be interpreted as the expected change in the money value of interest payments (interest credited). The expected real interest rate $r^*$ is determined, in accordance with (1.3), by the difference between nominal interest rate i and perfect, certain inflationary expectations $p^*$, adjusted by the inflation expectation factor $(1+p^*)$.

### 1.4. Results and implications

#### 1.4.1. *Time, scope and basic findings*

The theorem (1.1) is derived for a single period in which neither interest rates nor inflationary expectations change.[12] It contains no explicit indication of time and is accordingly static in character. Nor does the Fisher theorem differentiate expressly between different time horizons (short and long term). However, it is long-term in nature to the extent that, because

of the assumption of perfect foresight, it relies on a state of equilibrium, attainable in the long run, from a purely theoretical angle (see above). In such a situation of equilibrium, expected and observed variables coincide. As explicitly noted, it follows for all the expressions (1.1) to (1.3), because of the assumption of perfect foresight, that there is equality of expected and actual inflation rates ($p^* = p$). From this again, it follows that there is equality of real interest rates observed ex post and real interest rates expected ex ante, i.e. $r^* = r$.

The expression (1.1) also gives no indication of any restriction to apply it, for example, to a closed economy or, within a national economy, to a specific financial market (in the literature, reference is usually made to the market for long-term bonds). In point of fact, the theorem (1.1) is also applicable to (international) exchange markets.[13]

All the expressions (1.1) to (1.3) formulate, on the basis of the underlying assumptions, the "law" of translating *an interest rate from one standard into another.* In other words, they represent a formalization of Fisher's key sentence: "The rate of interest is always relative to the standard in which it is expressed.[14] None of the expressions (1.1) to (1.3) in itself determines an interest rate in any standard, whether a nominal interest rate or a real interest rate: "These rates are mutually connected, and our task has been merely to state the law of that connection. We have not attempted the bolder task of explaining the rates themselves.[15] For, as the derivation showed, the above formulae are consistently based on the assumption that the interest rate is already known in *one* standard. Thus, for example, the nominal interest rate in (1.2) can be determined only if the level of the real interest rate together with the level of inflationary expectations) is known a priori (for numerical examples, see Figure 1.1 below).

### 1.4.2. *Equation versus identity*

Under the basic assumptions postulated, the above formulations of the theorem are identities: they have the character of conditional equations which formulate the "law" of the connection between three variables – nominal interest rate, inflation rate and real interest rate. This fact has been of major importance in the practical application of the theorem, particularly the real interest theorem (1.3). Since the real interest rate is not directly observed statistically, the definition (1.3) provides the means of translating observed nominal interest rates into ex post real interest rates.

Nevertheless, the expressions (1.1) to (1.3) are not written as identities, but as equations. This is not an oversight, but is intentional, for Fisher's theorem would not be properly understood if it were characterized as a definitional relation. Together with the assumption of perfect foresight, a

certain type of *behaviour* is postulated: in a situation of equilibrium (with given ideal market conditions), market participants always adjust nominal interest rates perfectly to inflationary expectations (which are always correct). This behaviour linked to the expectation hypothesis, robs the theory of the quality of a true identity: it can be falsified if the expectation hypothesis, and hence the theoretically postulated adjustment of nominal interest rates to inflationary expectations should prove to be inconsistent with the facts. This question of the theoretical and actual adjustment of nominal interest rates is central to the literature which followed on from Fisher's work. This is also why references to the "Fisher relation" are usually references to the nominal interest theorem in the form of *equation* (1.2). In accordance with the double nature of the theorem, all the formulae (1.1) to (1.3) can be regarded both as definitions and as (empirically sound) behavioural equations; all possess the above-mentioned ambiguity.

### 1.4.3. *Restrictions and simplifications*

For the nominal interest rate i in (1.1) to (1.3), we have the restriction

$$i \geq 0.$$

The interest rate in a money which can be hoarded (in the future) without risk (which does not "spoil") cannot fall below zero: otherwise it would be preferable to hoard money than to lend it ex ante, that is to say consciously, at a loss. Expressed in interest theory terms, time preference for the present and the possibility of productive resource allocation generate a positive interest rate ex ante, given a constant or rising price level. There is another, and more important, restriction on the nominal interest rate, which has direct consequences for the real interest rate. As perfect foresight is assumed, the interest rate on money will never be smaller than any expected reduction in the purchasing power of money. That is, the rate of interest i is restricted to be equal to or larger than the expected inflation rate,

$$i \geq p^*.$$

In Fisher's terminology: The depreciation of money relative to goods cannot be greater than the interest rate expressed in the standard money.[16]

From the two restrictions it follows, according to the theorem, that the real interest rate is non-negative:

$$r^* \geq 0.$$

In summary: With perfect foresight, time preference (for the present) and possibilities of productive allocation of resources, market participants will not permit a negative expected real interest rate.[17]

The complete theorem (1.1) implies an indication of the spread between real and nominal interest rates. Where there are inflationary expectations, the real interest rate – cf. (1.3) – must obviously differ from the nominal interest rate by somewhat more than these expectations. In other words, the nominal interest rate – cf. (1.2) – must exceed the expected real interest rate by more than the expected inflation rate. This implies

$$i - r^* > p^* \text{ and } i > r^* + p^*.$$

Let us take a numerical example. Given a nominal interest rate of $i = 0.10$ (10%) and inflationary expectations of $p^* = 0.05$ (5%), we have, in accorance with (1.3), an expected real interest rate of $r^* = 0.045$ (4.5%). This figure is obviously different from the 5% produced by the popular,[18] simplified formula "real interest equals nominal interest minus inflation rate". The difference between the nominal interest rate and the expected real interest rate (5.5%) is greater than inflationary expectations (5%); and the sum of the expected real interest rate and inflationary expectations (9.5%) is smaller than the nominal interest rate (10%).[19]

In formal terms, the *simplified nominal interest theorem* is:

$$\boxed{i = r^* + p^* \quad p^* = p,} \tag{1.2'}$$

and accordingly the *simplified real interest theorem* is:

$$\boxed{r^* = i - p^* \quad p^* = p.} \tag{1.3'}$$

Both formulations can be regarded as correct only if continuous compound interest is assumed: in that case, the phenomenon which constitutes the difference between (1.2) and (1.2') does not occur.[20] The longer the time intervals at which interest is credited, and the higher the expected inflation rate, the clearer is the numerical discrepancy between the complete and simplified versions of the theorem, i.e. the more the spread between the nominal and real interest rates will differ from the value of inflationary expectations.

As already noted, the use of the simplified formulae (1.2') and (1.3') has become established in the literature. Only occasionally is reference made to the necessary assumption of continuous compound interest.[21] However, even then no mention is made of the fact that continuous compound interest payment virtually never occurs in practice.[22]

### 1.4.4. *The theoretical adjustment of nominal interest rates to inflationary expectations*

If inflationary expectations rise by $\Delta p^*$, nominal interest rates will immediately rise by the full amount of $\Delta p^* + \Delta p^*$ in accordance with the complete version of theorem (1.2), or by the amount, in principle lower, of $\Delta p^*$ in accordance with the simplified version of theorem (1.2'). Generally speaking, with a *given real interest rate*, rising inflation rates result in a proportional increase in nominal interest rates, and falling inflation rates in a proportional decrease in nominal interest rates; if there is no change in inflationary expectations, the nominal interest rate remains unaffected, all other things being equal. Where price stability is expected ($p^* - 0$), nominal and real interest rates coincide – a situation which Fisher precluded for the derivation of his theorem.[23] A continuously rising price level (i.e. a constant positive inflation rate $0 < p^* = \text{const.}$) is in theory associated not with continuously rising nominal interest rates, but with continuing high nominal interest rates. Under the terms of the theorem, the necessary adjustment of nominal interest rates occurs directly, completely and promptly.

This perfect theoretical relationship is illustrated in simplified form in Figure 1.1 for various hypothetical inflation rates and for a given "normal" real interest rate of 5%. Since here it is essentially to demonstrate the directions of movement and not the exact scales, it is sufficient in Figure 1.1 to take the simplified nominal interest theorem (1.2') as the basis.[24] The upper section of the diagram shows the assumed expected rates of change in the price level, roughly indicated by the slope of the line representing prices. The lower section shows the theoretical effects of various inflationary expectations on nominal interest rates.

The starting point of the curves plotted in Figure 1.1 is a situation of stable prices with $p^* = 0$ and consequently equality of nominal and real interest rates at 5%. If at the beginning of the first period there is a general expectation of a rise in the price level amounting to a constant annual 5%, the nominal interest rate "jumps" immediately to 10% according to Fisher's theorem and remains at this higher level so long as the price level is expected, with perfect foresight, to rise by a constant 5%. If in period 2 price stability is restored (at a higher level), the nominal interest rate will fall immediately to its initial level and remain there so long as there is price stability. It the expected price level rise by a constant 10% in the following periods 3 and 4, the nominal interest rate will rise immediately, at the beginning of period 3, by 10% to reach 15% and will remain there. If subsequently there are deflationary expectations (period 5), the nominal interest rate will fall immediately, to below the real interest rate; in the extreme case plotted, the nominal interest rate falls drastically, by 15 percentage points, to zero. The unchanged positive real

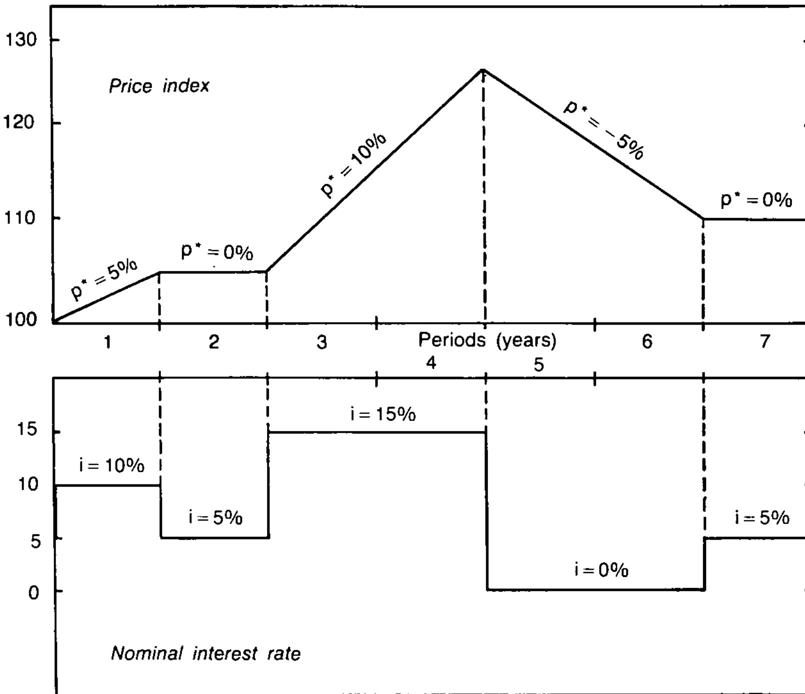



*Fig. 1.1* Theoretical, simplified relationship between inflationary expectations p* and nominal interest rate i assuming a constant real interest rate of 5%.

interest rate of 5% and the deflationary expectations also of 5% counterbalance one another in periods 5 and 6 in accordance with the simplified theorem.[25] Finally, in period 7, the nominal interest rate returns to its initial level of 5%, since there is price stability albeit at a higher level.

It is quite legitimate to describe the upper curve in Figure 1.1 alternatively as the expected or as the actual rate of price changes: since it is consistently assumed that there is perfect foresight, the expected and actual rates of price changes are the same, as shown above, i.e. $p^* = p$. In his explanatory comments, therefore, Fisher can afford for much of the time to refer only to observed changes in the price level.

As Figure 1.1 shows, a feature of Fisher's original theorem is that nominal interest rates react promptly and fully to prevailing inflationary or deflationary expectations. Consequently, where there are marked fluctuations in inflationary expectations, nominal interest rates will also show marked fluctuations. By contrast, the expected real interest rate typically

remains constant under the original theorem, that is to say it remains completely unaffected (or "neutral") by the perfectly anticipated, future rates of inflation. In other words, nominal interest rates adjust fully to price expectations: "If men had perfect foresight, they would adjust the money interest rate so as exactly to counterbalance or offset the effect of changes in the price level, thus causing the real interest rate to remain unchanged at the normal rate."[26] This "normal" real interest rate is evidently the interest rate obtaining in a situation of price stability (see Fig. 1.1), where the nominal interest rate and the real interest rate are at the same level.[27]

Lastly, a typical feature of Fisher's theorem, though one which is largely ignored in the literature, is the fact that the absolute spread between the (theoretically constant) real interest rate and the (theoretically fluctuating) nominal interest rate differs from $p^*$: rising inflation rates mean (assuming perfect foresight) that the absolute spread between nominal and real interest rates will become greater than the inflation rate $p^*$, since the product of $r^*p^*$ in (1.2) increases. Conversely, falling inflation rates will, through falling nominal interest rates, lead to a narrowing in the spread. This phenomenon is, as already stated, not reflected in Fig. 1.1. There, the basis used was rather the simplified version of the Fisher theorem (1.2'), so that the spread between the two interest rates is always precisely as great as the expected inflation or deflation rate.

## Annex: On Fisher's concept of interest

### A1. The agio concept: interest rates as intertemporal relative prices

Fisher uses Böhm-Bawerk's view of interest. Under this concept, the interest rate is seen as the premium (agio) on goods (assets) available in the present when exchanged for similar goods available in the future.[28] "Goods" may mean all kinds of durable assets, particularly money, but also a specific good or the (weighted) quantity of goods in a particular basket. The subject of Fisher's theorema is, as has been shown, the "translation" of a given percentage premium on money available in the present, i.e. a nominal interest rate, into the percentage premium on the quantity of goods available in the present, i.e. a real interest rate, assuming given inflationary expectations. Since the nature of the particular asset in question influences the premium, i.e. the interest rate, there are generally different interest rates for different assets; consequently, the interest rate must always be seen in relation to the asset or standard in which it is expressed. In accordance with the concept, therefore, there are any number of nominal and real interest rates in the sense that a percentage premium is paid on goods available in the present in exchange for goods

available only in the future: thus, for example, one can say that there is a (real) interest rate for non-monetary assets such as wheat, machinery, houses, etc. Similarly, there are any number of nominal interest rates if the premium on immediate availability relates to financial assets (money, securities). Variable periods, maturities, interest payment intervals or risk considerations increase the range of interest rates which may actually exist.

In microeconomic terms, on the basis of the premium concept of interest, an annual rate of interest characterizes the exchange relationship between an individual good $G_0$ available in the present ("today", "currently") and the same good $G_1$ available only in one year's time. (The indices denote the time of availability.) Today's present value price of good $G_0$ (for delivery now) is written as $P_{G0}$, and the present value price of the future good (for delivery at time 1) is written ad $P_{G1}$. If the availability of the present good is sacrificed for one year ($G_0$ is lent for one year), the right to $1+J$ future goods is acquired in exchange; J denotes the interest rate in accordance with the *definition*

(A 1.1.) $$1+J=P_{G0}/P_{G1}$$

or

(A 1.1') $$J=(P_{G0}/P_{G1})-1;\; P_{G0} > P_{G1}.$$

The definitional equations embody a general equilibrium approach, suitable for macroeconomic as well as microeconomic issues. The symbol J has been chosen to show that there is no determination of a nominal interest rate i or a real interest rate r. Depending on whether the good G denotes a physical asset or a financial asset, J is a real or a nominal interest rate.

The premium concept of interest very generally stresses only the time aspect of the interest phenomenon. Obviously, the interest rate can also be described here as an intertemporal relative price.[29] Terminologically, this is expressed by denoting $P_{G0}$ as the cash or spot price and $P_{G1}$ as the forward price.[30]

The identities (A 1.1) and (A 1.1') can also be interpreted as expressing the interest rate for any asset or good in the price of that good. Accordingly, J also denotes the *own-rate of interest* of a good.[31] The own rate of interest of money can therefore be called money interest and the own-rate of interest of a basket of goods as real interest, in accordance with Fisher's theory.

The concept of the own-rate of interest brings out particularly the basic idea underlying the theorem, i.e. that interest rates must always be seen in relation to the unit of goods or money in which they are expressed. The concept of the own-rate of interest stresses the relevant "standard", and

the premium concept of interest the underlying time aspect of the interest phenomenon. Both are entirely compatible with each other.

The formulation of the premium concept of interest in (A 1.1) is in line with the usual presentation. However, it is incomplete to the extent that it is based on an implied assumption, namely that the time interval $\Delta t$ for which an intertemporal relative price is expressed is irrelevant. in practice, this assumption may be justified in so far as interest rates are mostly annual interest rates, i.e. $\Delta t = 1$ year. Yet there are also exceptions, such as the setting of interest rates per month ($\Delta t = 1$ month) in the case of instalment loans.[32] It is therefore more correct to include the time element $\Delta t$ expressly in the formulation of the premium concept of interest. Accordingly,

(A 1.2) $$J = (P_{G0} - P_{G1})/(P_{G1} \cdot \Delta t); \; P_{G0} > P_{G1}$$

or implicitly

(A 1.2') $$P_{G0} = P_{G1} \cdot e^{\Delta t (1+J)} \text{ (continuous time).}$$

The obvious dependence of an interest rate on the reference period is reflected in the *dimension*[33] of the interest rate as an abstract number per time. This statement of dimension applies to all interest rates, irrespective of the standard in which they are expressed. For there is, it is true, an abundance of real interest rates, which, in line with the time premium concept (agio concept), are conceived of as the own-rate of interest for a quite specific non-monetary good or for a quite specific basket of goods – just like the distinction between nominal interest rates on financial assets in accordance with the nature of assets. However, Fisher's theorem cannot be taken as a formal expression of the dimension of real interest and nominal interest, even though the core substance of his theorem[34] has an apparent similarity with his general definition of dimension.[35]

Fisher's theorem is essentially suited to translating the interest rate from any unit of money into any unit of goods, and vice versa; as already stated, it formulates the "law" of the inner relation between two interest rates that are expressed in different units. In his empirical examination of the theorem (see Chapter 2 below), however, Fisher allowed this microanalytical aspect to recede into the background. From the abundance of physical assets, he takes a composite physical asset, namely a basket of goods, which then stands for the non-monetary, macro-economic asset "goods". This means that the focus is shifted to the movement in value, as specified in a given price index, of a macroeconomic sum of goods relative to money. Hence, the Fisher theorem should be assigned to the theory of money rather than to the theory of interest. From a practical point of view, however, it is quite appropriate to disaggregate the calculation of real interest using different price indices.

Overall, interest concept and dimension imply that the interest rate expressed in units of goods (in a particular basket of goods) does not show any particular characteristics as compared with any other interest rate expressed in units of money. Each theoretical interest formula too can basically be expressed in different standards, with the inner relation between the relevant interest rates being described by Fisher's theorem. For example, the Keynesian marginal efficiency of capital (an interest rate) could be expressed not only in the standard "money" but also in the standard "basket of goods for a representative price index": To quote Keynes: "If there were some composite commodity which could be regarded strictly speaking as representative, we could regard the rate of interest and the marginal efficiency of capital in terms of this commodity...".[36]

Conversely, a "marginal efficiency of money" can also be used to formulate a nominal interest (money interest) concept based on the time premium view. To quote Keynes once again: "Interest on money ... is simply the premium obtainable on current cash over deferred cash, so that it measures the marginal preference (for the community as a whole) for holding cash in hand over cash for deferred delivery. No one would pay this premium unless the possesion of cash served some purpose, i.e. had some efficiency. Thus we can conveniently say that interest on money measures the marginal efficiency of money measured in terms of itself as a unit".[37]

## A2. The concept of interest in terms of capital theory: the interest rate as a link between capital stock and income flow

The premium concept of the interest rate is independent of annuity, i.e. of the concept of a constant, annual payment flow (income flow) from capital. Because of this independence, the premium concept must as a matter of principle be distinguished from the capital theory concept of interest, which is based on the concept of annuity. The capital theory concept of interest focusses on the relationship between income Y, flowing from the use of a given amount of "eternal" capital assets K', and the value of these assets PK'. This relationship is defined, in respect of a constant income (net) flow of indeterminate (infinite) duration, as the capital theory concept of interest J', where

(A.1.3) $J' = Y/PK'$

PK' = value of the capital assets at a specified point in time

Y = income flow from the use of physical capital assets K' during a specified time interval

J' = capital theoretic rate of interest.

In accordance with the definition (A.1.3), the interest rate is simply the relationship between income and capital value. It could therefore also be described as the earnings price (user's price, income price) of capital.[38]

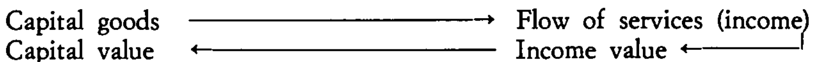
The distinction between the premium concept of interest and the capital theory concept of interest follows directly from the relevant underlying time element. Under the capital theory concept, the interest rate (J') applies for an infinite number of (annual) periods, because of the underlying concept of an annuity not limited in time. Under the premium concept, the interest rate (J) always applies only for the (annual) period or periods specified.[39] The concepts of interest are interchangeable only where there are numerical constancy in interest rates[40] and infinite time periods. Otherwise, the two concepts will each produce differing results.[41]

The value of capital assets[42] is, in accordance with (A.1.3), always related only to future, discounted (capitalized) income. The definition (A.1.3) therefore contains the general statement, in interest theory terms, that the value of the stock of capital assets lies solely in its value as a source of an income flow. This value is found by discounting the expected income flow to the present.

The present value of capital assets can be determined, in accordance with (A.1.3), from the value of the expected future flow of income from them; these is therefore the connection

$$\left\{\begin{array}{l}\text{Value of future income}\\ \text{from capital assets}\end{array}\right\} \xrightarrow[\text{(capitalization)}]{} \left\{\begin{array}{l}\text{Present value of}\\ \text{capital assets}\end{array}\right\}$$

In this relationship, the initial variable (income value) lies in the future and the result (capital value) in the present; the sequence accordingly runs counter to usual conceptions. The explanation for this lies in the distinction between physical entities and the values (prices) of these entities. Income in the sense of a physical flow of goods is, conversely to the relationship formulated above, the result of the use of physical capital goods: expenditure on goods ultimately always result from the combined remunerations of capital assets like human capital, physical capital, financial capital and land. A simple scheme representing both "directions of causation" might therefore look as follows:




This scheme devised by Fisher[43] shows that, for example, the wheat crop (real income) "naturally" depends on the land which yields it. But the *value* of the crop does not depend on the value of the land. On the contrary, the present value of the land depends on the value of the (expected future) crops. Therefore, in order to determine the present value

of the land, or of any other capital asset, two variables must be known: the future income from the asset and the interest rate by means of which the future income is translated into a present value.

Since the rate of interest is thus the bridge or link between a future income flow and a present capital asset, all income (applying a wide concept of capital) is ultimately to be interpreted as a stream flowing from the source "stock of capital assets". Accordingly, any analysis of income determination and development is necessarily confronted with the task of taking account of the stock-flow relationship, i.e. the above mentioned relationship between stock and flow variables. Consequently, conventional textbook approaches in which changes in the capital stock are not explicitly taken into account, inevitably ignore significant capital asset effects. Even if one defines capital assets more narrowly to mean "physical capital plus financial capital", the formation of physical capital (and hence the central role of investment in the economic process) forces us to take a stock-flow view. The Keynesian theory of investment was accordingly criticized on precisely this point at an early stage.[44]

Modern macroeconomic equilibrium theory has in recent decades, by almost universal consent (despite differences in detail), chosen the course of constructing models with stock and flow equilibria that influence one another mutually.[45] It may be said, though the statement may at first sight appear too presumptious, that modern macroeconomic theory has in the past two decades developed anew from a classical interest-theory nucleus (Wicksell),[46] namely from the time aspect that lies at the heart of interest rates (including the associated formation of expectations), from the necessary consideration of the interactions between stocks and flows and, lastly, from the analysis of the interplay of specific interest rates themselves.

The definition of the concept of interest set out in (A.1.3) leads ultimately to the traditional, classical distinction of the production factors "land", "labour" and "capital" and to the functional incomes "rent", "wages" and "interest" deriving from these production factors. Leaving aside differences of dimension between interest and rent,[47] the concept of interest (J') does not lead to classical "pure interest" in the sense of a pure income on the allocation of a specific factor referred to as physical capital. Rather, interest, on any production factor, is the relation between the price for the use of the factor and the price (value) of the source of those factor uses.[48]

In summary, it has to be borne in mind that, of the two basic ways of interpreting the interest rate, Fisher applied the time premium ("agio") concept for the purposes of his theorem. The literature which followed on from Fisher did not, pursuing his theorem, take up the distinction between that interpretation of the interest rate and the alternative interpretation in terms of capital theory. Although reference is made to the existence of two concepts of interest in modern writings on the theory of interest and the

theory of prices,[49] such references are made in isolation and are not linked up with the concept of interest used in Fisher's theorem. The particular appropriateness of the premium concept of interest to the time horizon underlying Fisher's theorem is not analysed in the literature, nor is account taken, even on a narrow view, of the numerical discrepancy that may arise from the two different concepts of interest.

## 2. Fisher's factual evidence

"No problem in economics has been more hotly debated than that of the various relations of price levels to interest rates. These problems are of such vital importance that I have gone to much trouble and expense to have such data as could be found compiled, compared, and analysed".[50] This, the opening sentence of the famous Chapter XIX entitled "The relation of interest to money and prices" in Fisher's *Theory of interest*, reveals the major importance which he attached to the empirical analysis of his theorem.

Fisher's initial concern was obviously an empirical test of his *original* theorem assuming perfect foresight. Because of the implied equality of observed and expected rates of price changes in the original theorem, we shold not be surprised that Fisher now refers to the (observed) price level and not expressly to price expectations. Moreover, prior to his empirical work, Fisher had emphasized, in his writings on monetary and interest theory, the existence and importance of expectations in the economy, however they might be formed. As early as 1896, for example, he wrote: "It is important to emphasize the broad fact that, in general, business foresight exists ... . It is the pratical man's business to foresee."[51] Even after several decades of continued work on questions relating to the theory of money, interest and business cycles, he made the following statement: "And today especially, foresight is clearer and more prevalent than ever before. The businessman makes a definite effort to look ahead not only as to his own particular business but as to general business conditions, including the trend of prices."[52] Today, more than half a century later, this statement can only be said to apply with greater force, given the large number of forecasts of all kinds which are available.

In view of these statements by Fisher and the equality $p = p^*$ assuming perfect foresight, it may be supposed that Fisher wanted to test his original theorem systematically, even if he started out by tailoring the problem to observed (and not expressly to expected) inflation rates.[53] Yet he realized from the outset that perfect foresight was a highly simplified assumption of price expectations, and one which he certainly did not a priori claim to be the case in practice. Contrary to what is maintained in the secondary literature,[54] Fisher did *not* expressly comment on the *way* in

which expectations are formed. We will see later whether it can be claimed that he attempted to quantify inflationary expectations with the help of a distributed lag model.

## 2.1. Comparisons of yields

As a first step in testing his theorem, Fisher concentrates on comparing the yields on bonds issued and quoted in two different standards.[55] He analyses two specific cases: fixed-interest US coin and currency bonds in the period 1870-96 and fixed-interest gold bonds and rupee (silver) bonds in the period 1865-1906. Fisher calculates in detail the actual ex post yields realized on the market in the relevant currencies and compares these interest rates with the ex post observed movement in the price of gold and in the exchange rate between rupee and sterling.[56]

The data show that yields on both bond issues were mostly different; changes in the relative value of the standards, in which interest rates were quoted, apparently played a role. Fisher sees the reason for the yield spread between the (otherwise largely homogeneous) bonds as being expectations, hopes and fears on the currency and security markes with regard to the relative trend in value of the different standards; in so doing, he expressly makes reference to his theorem. In 1870, for example, investors in US gold bonds realized 6.4%, while investors in US currency bonds were willing to accept a return of only 5.4%. Fisher puts forward the following explanation: paper had at that time depreciated significantly relative to gold in the United States, so that there was increasingly the expectation that it would in future rise again in value as compared with gold, perhaps even reaching parity with the hoped-for (or feared) resumption of the gold standard (official conversion of gold).[57] Fisher argues that, because it was expected that paper dollars would appreciate, the demand for currency bonds was relatively stronger than that for gold bonds, resulting in higher prices for and correspondingly lower yields on currency bonds relative to gold bonds.[58]

Fisher goes on to observe that after 1879, when there was a resumption of officially guaranteed equality in value between paper money and gold, the two bond rates remained very nearly equal for several years. It was not until there was talk of fundamental monetary changes (i.e. restoring free coinage of silver), which could have led to a preference for gold bonds, that the original yield differential was reversed; the yield on gold bonds fell below the yield on comparable bonds in the paper currency standard, in respect of which there was a fear of inflation.

It should be noted here, that in contrast to the (modern) adaptive expectation hypothesis Fisher evidently did *not* derive expectation formation solely from past price or exchange rate changes, but took other factors

into account as well. Fisher's expectation formation hypothesis implied here might accordingly best be described as "weak-form rational".[59]

In the comparison of yields on Indian gold bonds and silver ("rupee paper") bonds, the absolute differences in yields show a connection with *exchange-rate expectations* (as supposed by Fisher) of the Indian currency relative to the British gold currency.[60] In interpreting the observed yield spread, Fisher once again assumes the basic validity of his theorem in the sense that market fears of a decline in the exchange rate of the rupee relative to sterling are considered as the main explanatory factor.

Fisher expressly mentions, however, that the perfect foresight assumed in his theorem cannot exist in practice; rather, there are vague, underlying suppositions, hopes or fears as regards future yields and exchange rates: "Of course investors did not form perfectly definite estimates on the future fall, but the fear of a fall predominated in varying degrees over the hope of a rise."[61] He indirectly infers the formation of such expectations, with the similarity of the two bonds being a significant factor: "Inasmuch as the two bonds were issued by the same government, possessed the same degree of security, were quoted side by side in the same market, and were similar in all important respects except in the standard in which they are expressed, the results afford evidence that the fall of exchange (after it once began) was, to some extent, discounted in advance and affected the rates of interest in those standards."[62] Fisher therefore explicitly notes that there was *incomplete* adjustment of nominal interest rates.

All in all, Fisher's comparison of observed yields was meant to establish some first, rough evidence for his original theorem (in the simplified version); the role of expectation formation was expressly, albeit indirectly, examined.[63] Fisher intended to use the results of his yield comparisons only in very general terms – to support his basic thesis that an expected divergence in the relative values of two standards influences the relevant interest rates in these standards. It was not possible to determine an exact measure of this influence by means of yield comparisons; the evidence obtained was insufficient for a falsification of the theorem. Although on the basis of this evidence Fisher stated as early as 1896: "The adjustment of interest to price movements is inadequate",[64] he was more self-critical in his later publication *The Theory of interest* in which he stated that his attempts to quantify the influence of inflationary expectations on market interest rates on the basis of yield comparisons were methodologically unsatisfactory: "The preceding comparisons offer not exact measure of that influence ... Somewhat unsatisfactory attempts to do this ... were made in my *Appreciation and interest* but are not reproduced here."[65] In his later evaluation of his overall evidence, Fisher places greater emphasis on the results which he obtained using other statistical methods (see below).

## 2.2. Comparison of observed nominal and real interest rates

In the second stage of his empirical investigation, Fisher turns to a comparison of simultaneously observed money interest rates (nominal interest rates) and real interest rates.[66] On a superficial view, the comparisons of yields described above could also be fitted in here, for example by expressing the yield on gold bonds as the "real interest rate" and the yield on currency bonds as the "money interest rate" (which Fisher avoided doing). However, the relative change in the standard would then consist not in national price level movements, but would consist directly in (relative) exchange-rate movements. Ultimately, this would mean transposing or adapting the original Fisher theorem to explain the connection between international interest and exchange rates.

However, when nominal and real interest rates are compared, emphasis is generally placed on a contrast which specifically characterizes the focus of Fisher's theorem, namely the movement of nominal domestic interest rates in relation to the expected domestic rate of change in the price level, or, implicitly, in relation to expected real interest rates (expressed in the units of goods of a specified shopping basket). It is this distinction which leads to the typical question as to the influence of expected relative changes in the value of the (domestic) commodity price level – or of the (inverse) purchasing power of money – on nominal interest rates.

As regards Fisher's comparisons between nominal and real interest rates, the typical data problem arose: Fisher now could not resort to actually observed contracts under which interest rates in different standards (currencies) were agreed and quoted. He had to confine himself to comparing time series of nominal interest rates with deflated interest rates (i.e. interest rates adjusted ex post for prices changes): "All we can do is to note the changes in the price level, translate the actual rates in terms of money into real rates, and compare *successive* periods.[67] The comparison is therefore between observed nominal and real interest rates. The real interest rates are calculated in accordance with the original, simplified theorem: "The real interest rates are obtained by subtracting from the money rate for any period the rate annual change in the price level for the same period."[68]

Fisher uses two kinds of short-term nominal interest rates, namely "bank rates" and "market rates" in various national money markets[69] and in various periods.[70] He deflates these money market interest rates by the respective national wholesale price index.[71]

The periods investigated are conceived of as sub-periods of varying lengths (between three and twelve years) which together form a total observation period.[72] The sub-periods were each selected in such a way that they should show a significant change in the rate of price change as compared with the preceding and following sub-periods: they were not

selected with any reference to the results (i.e. whether these would show any consistency between the implications of the theory and the facts).

For each sub-period, Fisher calculates a (multi-year) average value for the nominal interest rate, the observed rate of price change and the observed real interest rate. For the period 1825-34 on the London market, for example, he calculates nominal interest rates of an average of 4.2% (bank rate) and 3.4% (market rate), an average annual 3% decline in the price level and, consequently, real interest rates of 7.2% (real discount rate) and 6.4% (real market rate on three-month bills).[73]

All in all, the second stage of Fisher's empirical procedure may be characterized as follows:

(i) The simplified original theorem, in version (1.2') and (1.3 ), is applied.

(ii) Calculated real interest rates are ex post observed rates.

(iii) The real interest rate figures are calculated for the average of several years; the simplified theorem is therefore applied over an extended time period.

(iv) The multi-year averages of the ex post observed real interest rates are based on deflation of quite short-term nominal interest rates, which means that multi-year averages of short-term real interest rates are calculated.

From the outset, Fisher himself was, from a macroeconomic point of view, critical of this procedure and pointed out that interest rates in the individual sub-periods were probably influenced not only by changes in the price level, but also, to widely varyng degrees, by other more fundamental factors: "Such comparisons are not very satisfactory, since no two periods ... differ only as to the state of the monetary standard ... Of course, influences other than changes in money affect interest rates."[74] An isolated, partial comparison of nominal and real interest rates ignores these other influences.

During the second stage of the investigations, the theoretical question is no longer stated expressly in terms of the role of *expected* changes in the price level. As in the introduction to Chapter XIX, Fisher refers only to "the" price level: "... support the theory that money interest rates move in the same direction as the price level."[75] Fisher's argument contains the following methodological and operational train of thought here:

(i) because the theorem assumes perfect foresight, the comparison of nominal and real interest rates *in one and the same period* does not require any explicit distinction between the observed and the expected rate of price changes ($p^* = p$).

(ii) If the factual evidence is inconsistent with the implications of the theorem (in particular if real interest rates are found to be variable), then the adjustment of nominal interest rates does not take place with perfect foresight.

(iii) If market participants are found to have imperfect foresight, the reasons for this must be sought. The question of how to modify accordingly the (original) theorem is left open.

Fisher indeed comes up with results which falsify[76] the implications of his theorem. Although he finds that there is an underlying positive link between money interest rates and the rates of change in the price level,[77] the adjustment of nominal interest rates to rates of inflation and deflation takes place only partially and slowly: "The adjustment is imperfect and irregular... When prices begin to rise, money interest is scarcely affected. It requires the *cumulative* effect of a long rise to produce a definite advance in the interest rate... The adjustment is very slow."[78]

This result calls into question the constancy of real interest rates implied in the original theorem: as a result of the slow and partial adjustment of nominal interest rates to the movement of prices, real interest rates (as calculated by Fisher) do not remain constant. They actually fluctuate to a significantly wider extent than nominal interest rates,[79] and indeed partly show negative values (which are theoretically impossible). Fisher rightly calls the fluctuations in real interest rates "erratic": for example, the real short-term rate of interest on the London money market fell from an average of + 1.2% in the period 1896-1913 to an incredible − 10.1% in the (inflationary) period 1914-20, only to soar again to + 15.7% in the subsequent (deflationary) period 1920-27.[80] Fisher summarizes, with reference to the theoretically required constancy of real interest rates: "What we actually find, however, is the reverse – a great unsteadiness in real interest when compared with money interest."[81]

This result is explained as being due to the discrepancy between theoretically perfect foresight and the imperfect foresight which actually exists among market participants: "... men are unable or unwilling to adjust at all accurately and promptly the money interest rates to changed price levels ..., there is very little direct and conscious adjustment through foresight."[82] Fisher attributes the inability of market participants to undertake promptly the theoretically "correct" adjustments in interest rates to widespread *money illusion*: "Most people are subject to what may be called "the money illusion" ... The erratic behaviour of real interest is evidently a trick played ... by the "money illusion" when contracts are made in unstable money."[83]

Fisher does not say anything more about the definition and origin of money illusion. However, we may suppose that Fisher had in mind the concept which he had put forward in general terms earlier, namely the

difficulty to perceive that the value of a monetary standard is constantly changing.[84] In other words, market participants are acting under money illusion when they are misled, despite the constancy of all real variables, to "adjust" real behaviour to price level changes.

Conversely, expressed in technical terms, there is freedom from money illusion when the supply and demand functions of market participants are homogeneous of degree zero in prices.[85] Accordingly, for the purposes of Fisher's theorem, money illusion means in quite specific terms that, despite changes in the price level and hence an unstable value of money (fluctuating purchasing power of money), market participants act *as if* the equation dollar = dollar remained unaffected in real terms: they react by "adjusting" the real interest rate rather than the nominal interest rate. If there were, by contrast, "freedom from money illusion", real interest rates would not react to fluctuations in the purchasing power of money.[86]

It should be noted here that Fisher later on considerably reduced the emphasis on money illusion as an explanation for his evidence, putting forward a *macroeconomic* explanation in terms of a specific transmission mechanism. Similar to earlier remarks,[87] he suggests that there is an *indirect* adjustment of nominal interest rates to inflationary expectations via business profits, volume of trade and the demand for loans: "The indirectness of the effect ... comes largely through the intermediate steps which affect business profits and the volume of trade, which in turn affect the demand for loans and the rate of interest."[88] Fisher also relies on business cycle considerations in explaining the extremely long time lags which he found for the (slow) process of interest adjustment (see below).

## 2.3. Correlation of nominal interest rates with lagged inflation rates

Fisher applied a third empirical test procedure in order to quantify more precisely the closeness of fit between nominal interest rates and inflation rates, including the implied speed of adjustment. He correlated nominal interest rates with lagged rates of change in the price level. It is above all this procedure that has been focused on in the subsequent discussion of Fisher's analysis and, as already indicated, has been interpreted as an attempt to quantify inflationary expectations.

### 2.3.1. *Focus*

Fisher's intention is to examine whether there is a positive link between nominal rates and rates of price changes: "... the theory being investigated is that interest rates move ... in the same direction as price changes".[89] There is no reference to the cause and effect relationship underlying Fisher's original theorem. Nevertheless, Fisher's formulation

of the question quite clearly falls within the overall context of empirically testing his theorem (and its implictions); the apparently vague statement of the problem in the above quotation may be said to be a correct statement of a correlation approach that does not comprise causality (but only a two-way cause and effect relationship). Given the assumption of perfect foresight, with $p^* = p$, one might again regard Fisher's statement as an operational formulation for testing the original nominal interest theorem. However, this would be to overlook two facts: firstly, the preceding tests (Sections 2.1 and 2.2) had led Fisher expressly to reject the original assumption of perfect foresight as being inconsistent waith facts and to reject the associated implication of perfectly adjusting nominal interest rates. Secondly, in his correlation estimates, Fisher replaces the earlier postulate of an immediate adjustment in interest rates by the concept of a lagged, gradual adjustment. The assumption of perfect foresight is, however, inconsistent with this concept. We must now examine what this means for the formulation of the theorem.

### 2.3.2. *Modification of hypothesis*

Fisher postulates a priori time lags, i.e. the existence of time intervals between price level changes and subsequent nominal interest changes. As an initial, preliminary test, he first questions the causality $p \rightarrow i$ implied in his original theorem, i.e. the effect of (expected) inflation rates on nominal interest rates: he also examines the reversed causality $i \rightarrow p$. In specific terms, correlation coefficients relating to discrete points in time are calculated for the relationship between observed rates of price changes and observed nominal interest rates, with Fisher assuming alternatively a lead or a lag of one to six years (in steps of one year) for the nominal interest rates. The data used consist of long time series for price changes and of long-term nominal interest rates (in the form of nominal bond yields) in the United States and Great Britain.[90]

The results,[91] whether with a lead or lag, show virtually no relationship between price changes and nominal interest rates: "These results suggest that no direct and consistent connection of any real significance exists ...".[92]

Fisher reacts to this outcome by introducing an additional statistical concept relating to the causality $p \rightarrow i$ (and only to this causality), borrowed from his earlier theoretical and empirical studies on the business cycle. It is the assumption that the postulated influence of price changes on nominal interest rates is *distributed in time*. The influence of price changes is therefore assumed not to exhaust itself in a single period, but to continue over several periods, with the intensity of the influence diminishing arithmetically as time passes by.

Assuming a distributed lag pattern, Fisher correlates current nominal interest rates with a weighted sum of past inflation rates. The latter is intended to measure length and intensity of the distributed influence of price changes on nominal interest rates. Fisher specifies the time profile of the distributed influence on the basis of arithmetically decreasing weights such that these taper off from the highest value at the beginning of the lag distribution to zero at the end of the distributed lag. The total number of periods n over which the influence is distributed is empirically determined, with successive extension of the (assumed) time horizon, according to the criterion of maximum correlation coefficient. Formally, the cumulative influence of past price changes is given as arithmetical average $\bar{p}$ at an assumed distributed lag over $l = 1, 2, \ldots$ n periods (years) and arithmetically decreasing weights $w_l$:

$$\bar{p} = \sum_{l=1}^{n} w_l \, p_{t-l}. \quad (2.1)$$

[93]

By correlating nominal interest rates with $\bar{p}$ for different lag figures n, the prompt adjustment of interest rates (which it was possible to assume in the first two test procedures on the basis of the original theorem) is from the outset replaced by the assumption of a time-consuming and gradual (i.e. lagged) process of adjustment of nominal interest rates to past inflation rates, among other relevant variables; i.e.

$$i_t = f(\ldots, \bar{p}, \ldots), \; f'(\bar{p}) > 0. \quad (2.2)$$

The dots in (2.2) indicate that Fisher did not in any way regard this relationship monocausally, but that he confined himself empirically to the analysis of *one* variable (i.e. $\bar{p}$) influencing nominal interest rates.[94]

It should be noted here that Fisher conceived of and carried out only correlation calculations and not regression calculations.[95] Equation (2.2) must not be misinterpreted in this way. The question as to what are the usual statistics and parameter values of Fisherian regression analyses[96] is accordingly understandable, but must remain unanswered.[97]

There is no mention of inflationary *expectations*: Fisher's underlying formulae (2.1) and (2.2) contain only ex post observable variables. His correlation calculations are to be understood not as the subject of an isolated sudy, but as the final empirical and, as will have to be shown later, also theoretical treatment of "interest and inflation", to which he devoted himself again and again in the course of more than three decades. An interpretation problem now arise here: the factual evidence already available to Fisher (Section 2.2) called into question his original theorem, pointing in particular to extremely imperfect or indeed non-existent price expectations, attributable to money illusion and/or time-consuming trans-

mission processes. Thereupon, Fisher carried out an empirical and statistical modification of his original 1896 theorem by introducing distributed lags (2.1). But he never expressly described the use of distributed lags as a method for an (approximate) quantification of inflationary expectations.

However, the literature which followed on from Fisher has designated *both together* (the modification and the expectation-theory interpretation) as "the" Fisher hypothesis or Fisher relation without ever stating clearly the interpretative character of such a secondary-literature assertion.[98] With reference to Fisher's original theorem it is simply assumed that $p = p^*$, and this "perfect foresight" implication gives, in conjunction with and in contradiction to the distributed lag (2.1), the *expectation formation hypothesis*

$$p_t^* = \sum_{l=1}^{n} w_l \, p_{t-l}. \tag{2.3}$$

(Symbols as previously.)

Hypothesis (2.3) has been unhesitatingly and rather unanimously praised as Fisher's great pioneering contribution to expectation theory, being a sort of forerunner of the modern variants of adaptive expectation formation, as formulated by Cagan and Friedman.[99] It was only rarely[100] pointed out that (2.3) differs distinctly from the implication of perfect foresight $p_t^* = p_t$;[101] for example, in the simplest case where $n = 1$, (2.3) yields $p_t^* = p_{t-1}$.

As regards justification of hypothesis (2.3), it may be pointed out (in the framework of a partial analysis confined to capital markets) that Fisher's distributed lag investigations were carried out in connection with an empirical test of his theorem. Fisher had already come to the conclusion that "perfect adjustment through foresight" could not exist, but he did not abandon his general emphasis on the role of expectations. This might indicate, together with his concluding interpretations (see Chapter 3 below), that he only changed the specific assumption on the way expectations are formed. However, Fisher was never explicit on this point. In addition, his *macroeconomic* prospective justifies serious doubts as to the usual interpretation of his theorem in terms of expectation theory (see Section 2.3.3 and Chapter 3 below).

If we posit the additional assumption $p^* = \bar{p}$, then the originally postulated dependence of nominal interest rates on *expected* price changes remains, and the *modified nominal interest theorem* (in its simplified version) can be formulated as

$$i_t = r_t^* + \sum_{l=1}^{n} w_l \, p_{t-l}. \tag{2.4}$$

with the "modern" interpretation that the summary term in (2.4) represents, together with (2.3), a hypothesis on the formation of inflationary expectations – a hypothesis which postulates a gradual, time-consuming, cumulative adjustment of nominal interest rates to these inflationary expectations. Because of this *new* thesis of only a gradual adjustment of $i_t$ to $p_t^*$, the original implication of a constant, expected real interest rate $r^*$ no longer applies; rather, it is implied that $r_t^* \neq$ const: as long as the rise in nominal interest ($+\Delta i$) lags behind an increase in inflationary expectations ($+\Delta p^*$), *the expected real interest rate $r^*$ must fluctuate inversely to inflationary expectations: rising inflationary expectations (where $\Delta i < p^*$) necessarily mean a decrease in the real interest rate ($-\Delta r^*$).* Within the framework of an expectation formation hypothesis like (2.3), we have – in respect of the (simplified) original theorem (1.2') – the implication

$$(2.5) \qquad (+\Delta p^* \rightarrow -\Delta r^* \mid +\Delta p^* > +\Delta i).$$ [102]

The *variability* of real interest rates is limited to that "transition period" within which the adjustment proces of i to p occurs. Consequently (2.3) and (2.4) describe the movements of the real interest rate in a situation of disequilibrium, in the transition period – in complete contrast to the original theorem assuming perfect foresight. And consequently (2.4) describes the "short-term" process of *imperfect* adjustment of nominal interest rates to price changes in the *disequilibrium transition period*, again in contrast to "long-term" equilibrium, when the adjustment process has been completed.[103] The real interest rate implication (2.5) was later reintroduced into modern writings, in a macroeconomic context and assuming perfect foresight, as the *Mundell effect*.[104]

The modification of the theorem in accordance with (2.4) implies as already indicated that the "neutrality" of expected price level changes is suspended in the transition period, and that expected real interest rates vary (inversely) with $p^*$. At the same time, this modified Fisher theorem is of far greater practical relevance, since it applies to precisely those (successive) situations of disequilibrium which characterize the practical reality of (cyclical) economic processes. All in all, this means that a central part of modern monetary theory, i.e. the neutrality of money, is banished into a politically irrelevant, distant and practically unattainable utopia in which equilibrium exists (see Section 1.1 above). Lastly, the modern debate on monetarism and its concept of a macroeconomic transmission mechanism are directly affected by this conclusion.

On view of such serious consequences, it is astonishing how unanimous the literature after Fisher has interpreted his theorem along the lines of (2.4), without weighting it against the original formula. Keynes is the only well-known author to have expressly drawn attention to the lack of clarity resulting from the contrast between Fisher's original theorem assuming

perfect foresight and Fisher's original theorem assuming perfect foresight and Fisher's use of distributed lags in the context of his (third) empirical test procedure: "It is difficult to make sense of this theory as stated, because it is not clear whether the change in the value of money is or is not assumed to be foreseen."[105] This criticism would not have been possible if Fisher had stuck to his original, explicit assumption of perfect foresight – or if he had expressly rejected the original theorem, in view of the test results, in favour of the modified theorem (2.4). Yet Fisher expressed himself clearly enough in business cycle investigations to allow us to understand and interpret his theorem (in the light of his factual evidence) as part of a disequilibrium analysis of macroeconomic transition periods; this is shown by the discussion of his results (below) on distributed lags (see Chapter 3).

### 2.3.3. *Results*

Fisher's concluding calculations using distributed lags for the above-mentioned American and British yearly data lead to striking results: the maximum correlation between nominal yield i and the weighted average of lagged inflation rates $\bar{p}$ is obtained for a distributed lag of altogether 28 years (mean lag:[106] about 9 years) with a correlation coefficient of almost 1 (+0.98) in the case of the British data (for the sub-period 1898-1924); and for a total distributed lag of 20 years (mean lag: about 7 years) with a correlation coefficient of +0.86 in the case of the US data.[107] As far as the closeness of fit is concerned, these results are in stark contrast to the previous results relating to *points* in time (see above). Fisher does not hesitate to accept the new results: "By assuming a distribution of effect of price changes over several years ..., the relationship between price changes and interest rates ... is clearly revealed. The high correlation coefficients ... show that the theory tested ... conforms closely to reality ...."[108] Fisher explains directly after this statement what theory it is that was being tested: "Our investigations thus corroborate convincingly the theory that a direct relation exists between P' and i [observed inflation rate and nominal interest rate, W.G.], the price change usually preceding and determining like changes in interest rates."[109] This is clearly the "theory" described in (2.1) and (2.2); once again, no mention is made of any expectational interpretation along the lines of (2.3).

Taking quarterly data for short-term interest rates (money market rates for 4 to 6-month commercial paper) and quarterly rates of price changes in the United States (total investigation period I/1890 to IV/1927)[110] Fisher essentially reaches the same results. Once again, correlations at certain points in time between rates of price change and leads or lags in nominal interest show little or no relationship, and once again the introduction of

distributed lags changes the picture drastically: for example, in the subperiod 1915-27, the correlation coefficient reaches a maximum value of +0.74 only when a total of 120 quarters (30 years) is included in the distributed-lag period – suggesting that a price change spreads its influence on nominal interest rates over no less than three decades.

From these results (regarding the influence of lagged price changes on nominal long-term bond yields and on short-term market interest rates) Fisher concludes that his basic idea of a positive relationship between nominal interest rates and inflation rates is independent of the length of period: "It would seem ... that price and interest fluctuations are governed by *one* law, not, as has been suggested, by two different opposing laws, for short and for long periods of time."[111] Evidently, Fisher himself here confuses the time horizon of his theorem with the term (or maturity) of the interest rates considered. It is correct that this "one law" applies in *principle* to short and long time periods – if it is understood as the expression of the general basic idea of the theorem, whereby each interest rate must be seen in relation to the standard in which it is expressed (see above). In specific, quantitative terms, however, the "law", by Fisher's own results, comes into operation only gradually, in the course of a long adjustment process.

## 3. Fisher's evaluation of evidence

### 3.1. Explanation of the Gibson paradox

The time series used by Fisher for the nominal yields of British long-term securities (consols) was derived from investigations carried out by A. H. Gibson, who earlier established a close positive correlation between the price *level* and nominal interest rates (specifically, nominal yields on consols).[112] Keynes termed this empirical phenomenon the "Gibson paradox"; he regarded it as a paradox because it contradicted classical interest theory under which "the" interest rate, though determined by fundamental real factors such as saving (or time preference) and capital productivity (or opportunity to invest), is independent of "monetary" factors such as the quantity of money and the price level.[113]

At the same time as Keynes, Fisher was also examining the empirical findings presented by Gibson. Fisher's main concern was to make clear empirically and theoretically the connection or difference between the Gibson paradox and his own theorem. His starting point was the basic distinction between the price level and the rate of price change (inflation rate), i.e. between the absolute level of a (price) series and its first derivation: the Gibson paradox relates to the observed (positive) relationship between nominal interest rates and the *price level*; the Fisher theorem

(in whatever formulation) relates to the (positive) relationship between nominal interest rates and *rates of change* in the price level (inflation rates). Fisher does not stress the fact that his theorem relates to *expected* inflation rates, as these can be linked up theoretically and empirically with observed inflation rates only if an expectation formation hypothesis is introduced; instead, he argues throughout on the basis of observed rates of change in the price level.

Fisher first reproduces the facts of the Gibson paradox by calculating coefficients for the long series of British and American yearly bond yields and price indices (see above). The correlation coefficients are highest for a lag of about one year in nominal interest rates. The values of the correlation coefficients fluctuate, with one exception, around +0.9 and consequently show a very close positive relationship in accordance with the Gibson paradox. This factual evidence is confirmed with quarterly data and short-term interest rates. Fisher summarizes as follows: "These highly significant correlations seem to establish definitely that over long periods of time high or low interest rates follow high or low prices by about one year."[114]

Fisher also checks this result statistically against possible trend influences. De-trended series (consisting only of cyclical, seasonal and irregular components) show similarly high correlation coefficients between 0.7 and 0.8, when interest rates are lagged by one year.[115] In periods of marked fluctuations in the price level, the relationship with nominal interest rates is particularly close if the short lag of one year is applied.

Fisher summarizes as follows: "It is quite definitely demonstrated that ... the effects of price movements are felt rather quickly upon the rates of interest, even in the case of long-term bond yields."[116]

The empirical evidence of the Gibson paradox whereby high nominal interest rates are accompanied by high price levels and low nominal interest rates by low price levels is explained by Fisher as an *accidental consequence*[117] of the empirical relationships that he was able to determine in investigating his theorem; these empirical relationships were:

(1) Nominal interest rates tend to be high when the price level is rising (inflation) and low when the price level is falling (deflation).

(2) There is a time lag in the adjustment of nominal interest rates to rates of price change, with the adjustment being distributed over a period of almost three decades.

From a theoretical point of view, Fisher equates the long period referred to in (2) (defined by the distributed lag corresponding to a maximum correlation coefficient) with his concept of a transition period; he is evidently resorting here to a basic reasoning which he developed earlier in analysing the quantity theory of money. The transition period is

therefore (despite its length) denoted as a "short-term" period of adjustment, in contrast to a "long-term" state of equilibrium once the adjustment processes have been completed, including real interest-rate adjustments.[118]

In the transition period, there is a cumulative adjustment of nominal interest rates to (expected) rates of price change in accordance with Fisher's (technical) distributed lag assumption. If a large part of the cumulation has taken place towards the end of the transition period, then precisely because of the adjustment process, a high interest level will correspond to a high price level. However, Fisher argues that this phenomenon, i.e. the Gibson paradox, would undoubtedly disappear if the high price level were to remain unchanged: for, given a zero rate of price change, i.e. a constant, albeit high price level, nominal interest must according to Fisher's theorem fall again to the level of the real interest rate. The chart in Chapter 1 (see page 15) demonstrates this clearly.

In fact, however, the price level is found not to be stable over time (e.g. in the long periods observed by Fisher and Gibson for Great Britain and the Unites States), but to be constantly changing. Consequently, there is literally *no time* for nominal interest rates to react fully, in accordance with Fisher's theorem, and to fall to the level of the real interest rate in any situation of price stability.

The Gibson paradox is thus attributed to the combined effect of the lack of price stability and the adjustment dynamics of the modified (simplified) Fisher theorem. Fisher summarizes: "Thus, at the peak of prices, [nominal, W.G.] interest is high, not because the price level is high, but because it *has been rising* and, at the valley of prices, interest is low, not because the price level is low, but because it *has been falling.*"[119]

The empirical relation described by the Gibson paradox is therefore an "accidental consequence" of the Fisher theorem to the extent that, without the ("accidental") fluctuations in the price level, i.e. lack of price stability, it could not occur in the long run, i.e. in a situation of equilibrium. Fisher thus clearly bases his argument on the classical dichotomy: he quite explicitly regards it as inconceivable that, in a state of long-term equilibrium, a high price level, resulting for example from preceding monetary expansion, should be accompanied by a higher nominal interest level than that (lower) interest level which had obtained *before* the inflationary rise in prices: "The price level as such can evidently have no permanent influence on the [nominal, W.G.] rate of interest except as a matter of transition, from one level or plateau to another."[120] In a situation of long-term equilibrium, therefore, the price *level* is *neutral* in relation to nominal interest rates.

After the Second World, the Gibson paradox has been "rediscovered" as a result of persistent inflation problems facing the Western industrialized countries.[121] In connection with empirical discussions of the more recent documentation on the long-term relationship between price level

and nominal interest rates, criticism was expressed of Fisher's explanation of the phenomenon.[122] We will not enter into this debate in detail.

However, a general point should be made: a number of authors have obviously not taken note, or have not taken full note, of Fisher's theoretical and empirical arguments as described above: emphasizing the extremely long time lags detected by Fisher, they interpret the (modified) Fisher theorem *itself* as a reformulation of the Gibson paradox. In fact, in purely empirical terms, an average of rates of price change over a number of decades is an approximative indicator of the price level itself. Consequently, under this interpretation, the Fisher theorem and the Gibson paradox are regarded virtually as synonymous expressions.[123] In view of Fisher's theoretical and empirical arguments outlined above, this identification of the two as equivalent is incorrect, since it fails to give a true reflection of Fisher's point of view; equating the Gibson paradox with the Fisher theorem is a misinterpretation. This can be seen most clearly if one assumes a stable price level; in such circumstances, Fisher's theorem and the Gibson paradox are *not* consistent with each other (cf. Chapter 1 and Chart 1.1).

### 3.2. Long lags and the transmission mechanism

Even to Fisher, it seemed fantastic, at first glance, to ascribe to price changes which occurred two or even three decades ago any influence affecting the rate of interest today.[124] He nevertheless tried to find precise economic reasons for this long lag phenomenon.[125] He first draws analogies on the long-term economic effects of natural disasters to substantiate the phenomenon of an extremely long period of impact as such. Thereafter, Fisher tries to explain his long lags within the framework of business cycle theory, emphasizing theory and evidence of the transmission of price changes to nominal interest rates. The deeper significance of Fisher's analysis here lies in the theoretical incorporation of the modified theorem into a structural, macroeconomic disequilibrium model. Fisher's discovery of the "Phillips curve" should also be noted in this connection (see below).

At a very early stage of his research, Fisher envisaged a causal relationship running from price rises via increased macroeconomic activity (more trade, greater demand for capital goods) and hence increased demand for money and credit, to higher nominal interest rates.[126] Later, he refined his view on the business cycle effect by including profit expectations of businessmen: "Rising prices increase profits both actual and prospective, and so the profit taker expands his business. His expanding or rising income stream requires financing and increases the demand for loans."[127] Fisher's argument here must be seen as a "concentrate" of his previous analyses of the transmission process in the context of the quantity theory

of money.[128] In his *Purchasing power of money*, he presents a dynamic analysis of nominal interest adjustment in the disequilibrium of the transition period from one (theoretical) state of equilibrium to the next.[129] Fisher describes a *monetary business cycle theory* of Wicksellian provenance, the main item of which is the effects of discrepancies between a "normal" or "natural" interest rate and the nominal interest rate; in so doing, he equates his real interest rate with this normal interest rate. The discrepancies result from money illusion and lead amongst other things to cyclical changes in credit demand. When the role of the banks is included, the analysis produces a sequence which Fisher summarizes as follows:

"1. Prices rise ...
2. The rate of interest rises, but not sufficiently.
3. Enterprises, encouraged by large profits, expand their loans.
4. Deposit currency ... expands relatively to money [notes and coins, W.G.].
5. Prices continue to rise ...."[130]

An initial price rise therefore triggers a chain reaction which tends to repeat itself. This thesis, which has regained relevance today, is based on the inadequate adjustment of nominal interest rates to the "normal" or "natural" rate: "Rise of prices generates rise of prices, and continues to do as long *as the interest rate lags behind its normal figure*."[131]

The significance of this statement by Fisher (who might be described as a quantity theoretician here) would be only incompletely grasped if it were seen solely, or primarily, as an argument to justify his long lag evidence. Rather, the transmission concept just sketched is to be understood as a theoretical basis for the modified Fisher theorem itself, i.e. as a disequilibrium framework incorporating precisely those fluctuations in real interest rates which Fisher diagnosed as a consequence of the gradual and incomplete adjustment of nominal interest rates. We must not be misled here by Fisher's own evident confusion of real interest and interest on capital, which can be assumed to be identical only in a situation of equilibrium (e.g. as in the original theorem): it must be remembered that the monetary business cycle theory adduced means that a precise distinction must be made between Fisher's original and modified theorems.

It is further evident that, while Fisher's expectations do play a role in the cyclical model described, it is not inflationary expectations, but firms' profit expectations which are the relevant expectation factor (see above). Although the modified nominal interest theorem is incompatible with this, the way in which it interprets expectations is not. Some signs that this is coming to be realized have begun to emerge recently in the literature. For example, B. Friedman makes the following statement: "Somewhat astonishing to the modern reader, Fisher's suggested interpretation followed

Knut Wicksell in noting that higher prices usually meant a greater nominal volume of trade, which in turn increased the demand for money, and hence increased nominal interest rates for given bank reserves. What is surprising about this interpretation is that, as rendered by Fisher, it has nothing whatever to do with [price, W.G.] expectations."[132]

The "Phillips curve" relationship which is built into the cyclical framework as postulated by Fisher[133] results, according to Fisher, in a temporary (or, in modern terms, transitory) decrease in unemployment if (observed) inflation rates are rising: "When the price level is rising, a businessman finds his receipts rising as fast, on the average, as this general rise of prices, but not his expenses ... The businessman, therefore, finds that his profits increase ... Employment is then stimulated ... for a time at least."[134] The *short-term* inverse relationship betwen inflation and unemployment, which has recently been "rediscovered" in a series of investigations and was examined in connection with Friedman's "natural rate hypothesis",[135] therefore complements the described cyclical effects by affecting business profits.

Fisher's thinking on transmission theory cannot be pursued further in detail here. The business cycle framework and the Fisher-Phillips curve cannot, as qualitative considerations, be taken to indicate any specific (long) period of duration of interest adjustment processes. Fisher therefore supplemented his theoretical considerations on the (macroeconomic) transmission process with quantitative results taken from earlier investigations.[136]

According to these, the lag with which the volume of trade (in present-day terms, an approximation of total gross national product) affects price changes is distributed over about two years. By adding the evidence (at Fisher's time) of a lag of nominal interest rates behind the volume of trade amounting to some 14 months, Fisher only obtained a combined lag between inflation and the nominal interest rate responses of just over three years.[137]

The extremely long distributed lags calculated by Fisher could not therefore be substantiated through evidence in transmission theory terms. Fisher simply noted this failure and left it at that. The inconsistency between Fisher's distributed lag calculations, which he interpreted as a quasi "reduced form" of his monetary business framework, and the structural transmission lag results remains an unresolved issue.

### 3.3. Reversed causality and interest-rate policy

In his explanation of the Gibson paradox, as in the context of his busines cycle theory, Fisher took account of the role of the banking system and the "reversed" causality: nominal interest rate → price level. In any

inflation process, the question arises of how far the banking system is willing and able to finance it. The scope for appropriate monetary expansion (e.g. loan expansion by credit institutions) is generally restricted by institutional conditions and central bank policy. During Fisher's time, monetary rules governing gold currency were among the factors limiting the expansion potential of a (national) banking system including the central bank.

if at some stage during a continuing inflationary process, national monetary expansion comes up against a quantitative limit, then, according to Fisher, excess demand arises on the money and security markets, which the banks resist by further raising nominal interest rates. The rise in interest rates will ultimately reach a point where contractive quantity and price effects are triggered in the economy.

The reversed causality: nominal interest rate → price level therefore depends on explicit account being taken of nominal interest rate policy of banks. Accordingly, Fisher's causal relationship: higher inflation → higher nominal interest rates is supplemented by the further consideration that high nominal interest rates are initially accompanied by a rising price level (Fisher theorem) and then by a falling price level (credit restriction, reversed causality with reversal of the algebraic sign), and vice versa. Consequently, nominal interest rates are high before, during and after the "peak of prices" and low before, during and after the "valley of prices".

Hence we have a positive correlation between nominal interest rates and the price *level*, i.e. the Gibson paradox.[138] It should be noted that Fisher's argument here is relevant from a present-day monetary point of view, for example with regard to the Bundesbank policy of high interest rates in the period 1979-81.

We can see from this reasoning that Fisher clearly accepts the thesis of a *mutual* relationship between nominal interest rates and inflation: *both* directions of causation must therefore be taken into account analytically.[139] In addition, he accepts the causality between nominal interest rates and inflation as *the* real focus of any central bank policy aimed at stability: according to Fisher, an (exogenous) increase in nominal interest rates by the central bank will always tend to pull down the price level, and vice versa in the case of a decrease in nominal interest rates.[140] He states that this inverse causal relationship between nominal interest rates and the price level is a fact which has been quite well established and that it is made use of by central banks in formulating their banking and credit policies.[141] Quite clearly, this is once again a position which is relevant with regard to the modern controversy surrounding "money supply" theory. One can justifiably argue that Fisher anticipated by 50 years the basic idea of money stock control through interest rates, as explicitly applied, for example, by the Deutsche Bundesbank[142] and the British monetary authorities[143] in recent years.[144]

The mutual causation between nominal interest rates and rates of price change does not represent any inconsistency in Fisher's view. In support of his argument, he refers to the different time horizons involved in the various directions of influence: in contrast to the very long lags of influence with regard to his modified theorem, he points to a very short lag in the reversed direction of causality, nominal interest rates → inflation: a decrease in nominal interest rates for example, will initially produce a rapid increase in economic activity and in prices. After a few months, increased prices will have a feedback effect on nominal interest and will pull the interest rate up.[145] Quite clearly, Fisher's analysis here again anticipates a modern monetarist's proposition, namely Friedman's rejection of nominal interest rates as appropriate "indicators" of central bank policy.[146]

The present-day relevance of Fisher's discussion and application of his modified theorem is not confined to his general references to the interactions between nominal interest rates and rates of price change. Through detailed consideration of the conditions governing monetary expansion by banks, particularly of the *reserve position of credit insitutions* and its importance for money market conditions, Fisher has something specific to say on the role of bank in the monetary expansion process.[147]

His point of departure is the practical observation of an inverse relationship (significant negative correlation) between the level of bank reserves and (short-term) money market rates.[148] From this, in view of national and international liquidity flows, he concludes that there is quite a powerful economic influence exerted by the "banking machinery",[149] which may quite considerably interfere with the "normal" economic transmission mechanism. In particular, the way in which banks behave (i.e., in more modern and more precise terms, their portfolio behaviour) is *largely responsible for business cycles* – an argument to which Fisher returned again and again.[150] His proposition that the banking system has a potentially destabilizing influence corresponds to a basic tenet of modern monetarism.[151] Fisher traces this proposition back to the chain of causation: bank reserves → nominal interest rates → business cycle → inflation. If the "interlink" nominal interest rates → money stock development is also inserted, and if one "rounds off" with Fisher's theorem as a feedback relationship inflation → nominal interest rates, one obtains a self-contained sequence of effects which is consistent with modern macroeconomic theory.[152]

Lastly, from Fisher's consideration of the relationship between bank reserves and (short-term) nominal interest rates, we can make a distinction with regard to the relation between money stock and nominal interest rates: the relationship bank reserves → money market interest rates must not be confused with the very different, and generally incorrect, assertion that money interest rates are high when money is scarce, and low when

there is an abundant quantity of money. The assertion is incorrect because, according to Fisher, the nominal interest rate is *not* to be understood as the "price of money".[153]

The conceptual discussion of the rate of interest as price of money, price paid for the use of money, marginal efficiency of money or purchasing power of money (as inverse of the price level) is, however, misleading in the present context.[154] Rather, Fisher's position points to his earlier findings on the quantity-theory relationship betwen money stock and price level: during long transition periods, an abundant quantity of money does not necessarily raise the price level proportionately (if at all). Hence, monetary expansion, inflationary developments, and reactions of nominal interest rates are quite loosely connected phenomena in transition periods. Therefore, in reality, we may witness monetary expansion without significant interest rate effects in the "short run".

On the basis of the theory of interest, one can at most speak of a relationship between a relative price (i.e. the interest rate) and the purchasing power of money.[155]

### 3.4. Practical application

On the basis of a scenario of incomplete, time-consuming and indirect adjustment of nominal interest rates to inflation or inflationary expectations, accompanied during long transition periods by imperfect foresight, money illusion and marked inverse fluctuations in real interest rates,[156] Fisher draws conclusions on the applicability of his original theorem in theory and practice: since, for example, when the price level is rising, interest rates increase slightly in nominal terms, but fall sharply in real terms (and vice versa when the price level is falling), it is *in practice* of the greatest importance to have regard to movements in the price level. This is because the very imperfect adjustment of nominal interest rates to (expected) inflation rates leads, in the case of financial assets with a fixed nominal interest rate, to an unforeseen loss for lenders and an unforeseen gain for borrowers: "It is consequently of the utmost importance, in interpreting the rate of interest statistically, to ascertain in each case in which direction the monetary standard is moving and to remember that the direction in which the interest rate apparently moves is generally precisely opposite to that in which it really moves.[157]

Fisher thus addresses the practical businessman. He refers him to the currently observed rates of price change as the main basis for distinguishing true, i.e. real, from apparent, i.e. nominal, interest movements.[158] In view of the obvious falsification of his original theorem because of the ultimate imperfectness of expectations and foresight, Fisher logically refrains from any extensive theoretical evaluation of his theorem and instead

places the emphasis on its practical relevance. "If the money rate of interest were perfectly adjusted to changes in the purchasing power of money – which means, in effect, if those changes were perfectly and universally foreseen – the relation of the rate of interest to those changes would have no practical importance but only a theoretical importance. As matters are, however, in view of almost universal lack of foresight, the relation has greater practical than theoretical importance.[159] The literature has not followed Fisher in this approach. Rather, in partial and macroeconomic analyses, the theoretical aspects have been emphasized. This might be justified, at least partially, by the steady disappearance of "money illusion" in recent years. Nevertheless, detailed practical investigations into the calculation and application of real interest rates at financial markets are still urgently needed.

## Summary

1. Fisher's original theorem in complete form is as follows:

$$(1+i)=(1+r^*)\,(1+p^*)$$

$i$ = nominal interest rate,
$r^*$ = expected real interest rate,
$p^*$ = expected inflation rate (rate of price change).

It is based amongst other things on the assumption of perfect foresight. From this it follows that there is equality of expected and actual variables, i.e.

$$p^*=p$$

and

$$r^*=r.$$

The theorem can be considered as a static equilibrium condition. It does not a priori contain any restrictions as to its scope. By multiplying out and transforming, we obtain the *original nominal interest theorem*

$$i=r^*+p^*+r^*p^*$$

and the *original real interest theorem*

$$r^*=(i-p^*)\,/\,(1+p^*).$$

All the formulations are interpretable as equilibrium definitions *and* as empirically testable behavioural equations, with the restrictions

$$i,\ r^{*} \geq O,$$

such that, by assumption,

$$i \geq p^{*}.$$

When inflationary expectations change, the spread between the nominal interest rate and the real interest rate does not remain equal to inflationary expectations; instead, we have

$$i - r^{*} > p^{*}.$$

The reason for this is, technically speaking, the absence of continuous compounding (i.e. discontinuous time). In terms of economic interpretation, the inequality arises due to the forces behind changing inflationary expectations.

Anyway, the nominal interest rate must exceed the expected real interest rate by *more than* the expected inflation rate. We can delete "more than", in the *simplified versions* of the original nominal interest theorem

$$i = r^{*} + p^{*}$$

and of the original real interest theorem

$$r^{*} = i - p^{*}.$$

The simplified versions ignore expected changes in the purchasing power of interest payments, i.e. the corective terms $r^{*}p^{*}$ and $(1 + p^{*})$ in the above, complete formulations. The simplification yields a precise statement in the case of continuous, infinitely frequent interest payments (compound interest). If the corrective terms $r^{*}p^{*}$ and $(1 + p^{*})$ are close to zero and close to 1 respectively, the simplification may be justified empirically.

Because of the assumption of perfect foresight, the original theorem in all its versions implies a *prompt and complete adjustment of nominal interest rates* to inflationary expectations. Hence, in theory, expected real interest rates are *constant*: they are not influenced by (changes in) inflationary expectations ("neutrality", i.e. homogeneity of degree zero in prices).

The original theorem formulates in general terms the "law" of the mutual interrelationship between (two) interest rates which are expressed in different standards. It does not in any way claim to explain the interest

rates themselves, whatever the standard in which they are expressed. Fisher dealt with an explanation of interest rates in his interest theory proper – on the assumption of a constant price level i.e. equality of nominal and real interest rates.

2. Fisher proceeds in three stages in carrying out an *empirical investigation of his theorem*. In the first two stages (comparison of yields and comparison between nominal and real interest rates calculated ex post), Fisher tests the simplified versions of his original theorem.

Fisher's evidence (on long-term bond yields and short-term money market and bank rates) suggest that nominal interest rates and expected inflation rates broadly move in the same direction. However, nominal interest rates adjust only slowly, incompletely and indirectly to expected inflation rates. Real interest rates are consequently not constant, but fluctuate, and indeed actually fluctuate to a far greater extent than nominal interest rates.

Fisher attempts to explain the factual evidence firstly on the basis of the existence of *money illusion*, which prevents perfect anticipation of inflation (prevents "perfect foresight"). Secondly, he puts "indirectness" down as a dynamic macroeconomic transmission mechanism: the chain of events from inflation → expected business → employment and overall economic activity (Phillips curve!) →loan demand → nominal interest rates explains "indirectness" and suggests a time-consuming pattern of adjustment.

In response to the factual evidence which contradicts his theorem, Fisher concentrates lastly on quantifying as precisely as possible the relation between nominal interest rates and observed inflation rates and the time lag involved. For this purpose, as the third stage in his empirical investigation, he calculates correlation coefficients to determine the closeness of link between nominal interest rates and distributed lags in inflation rates. The original assumption of perfect foresight together with the associated implication of perfect nominal interest adjustment is dropped. Specifically, Fisher tests the relationship between the present nominal interest rate $i_t$ and the weighted average $\bar{p}$ of the distributed lags of inflation rates $p_{t-l}$, i.e. $\bar{p} \rightarrow i$, with arithemetically declining weights $w_l$, i.e.

$$\bar{p} \rightarrow i, \bar{p} = \sum_{l=1}^{n} w_l \, p_{t-l}.$$

Lagged past inflation rates show a very close, positive correlation with current nominal interest rates, with the lags extending *over two to three decades*. These results explain and quantify, in an extreme way, the

evidence found earlier that nominal interest rates do not adjust perfectly to inflation rates.

3. Fisher's correlation cannot be regarded as an empirical investigation of his original theorem. Rather, in view of the apparent lack of foresight found earlier, it serves only to test the basic idea of the theorem: that interest rates are always relative to the standard in which they are expressed. Consequently, Fisher tests specifically the relation between money interest rates and *observed* lagged changes in the purchasing power of money (or in the price level). In a macroeconomic disequilibrium discussion, Fisher described this relationship in terms of the "transition period" of an economy from one state of equilibrium to another. Nowhere did Fisher expressly designate his distributed lag approach as an approximation for inflationary expectations.

The *modified nominal interest theorem* which Fisher analysed using correlation methods is as follows:

$$i_t = r_t + \sum_{l=1}^{n} w_l \, p_{t-l}.$$

(Symbols as above; $r_t$ is the observed real interest rate.)

In the secondary literature, the summary term has been interpreted as *Fisher's* hypothesis of a gradual, adaptive formation of inflationary expectations, i.e.

$$p_t^* = \bar{p} = \sum_{l=1}^{n} w_l \, p_{t-l}.$$

The justification of this interpretation must be questioned in view of Fisher's macroeconomic transmission mechanism for the "transition period", which emphasized not inflationary, but *profit* expectations (of businessmen).

The modified version is a dynamic formulation. In the "short term", i.e. for the duration of a transition period covering roughly a decade, it implies an incomplete, slow, and cumulative adjustment of nominal interest rates to past rates of price change. Consequently, in such periods a (fairly considerable) *variability in real interest rates* is implied: real interest rates fluctuate inversely to the change in inflation rates. Since fairly long historical phases of price stability (rates of price change = zero) cannot be observed, we have in practice a series of transition periods following and overlapping one another. Theoretically speaking, there exists a sequence of (price) disequilibria: the modified nominal interest theorem denotes practi-

cally a *permanent state* of incomplete and lagged adjustment of nominal interest rates to observed rates of price change.

4. The parallel (positive) causation: inflationary expectations → nominal interest rates is consistent with the reversed causality of a Wicksellian relationship: nominal interest rates → inflation. *"Fisher" is consistent with "Wicksell"*. Fisher expressly emphasized the interactions between "interest and inflation". However, he estimates that there is a considerably shorter time lag for the causality interest rates → inflation than for the reversed causality within the framework of his modified theorem.

Fisher's explanation of the *Gibson paradox* is based on the evidence which the modified nominal interest theorem suggests for "short-term" transition periods plus the complementary proposition of interest rates affecting business and prices: Given a variable price level, the Gibson paradox results from the interaction between slow and incomplete adjustments of nominal interest rates to inflation and the fluctuating price level itself. Where there is long-term stability in the price level, the (modified) nominal interest theorem and the Gibson paradox are not compatible with one another.

The modified nominal interest theorem is consistent with quantity theory statements on the long-term determination of the price level by the quantity of money only on the basis of the "classical dichotomy" between monetary and "real" economic phenomena. However, "neutrality" of monetary changes in relation to physical "real" economic magnitudes applies without restriction only to the original Fisher theorem, which implies a corresponding constancy of real interest rates. The modified nominal interest theorem is in line with the classical postulate of neutrality only in a *situation of long-term equilibrium*. In fact, however, that long-term situation of equilibrium in which complete adjustment of nominal interest rates to changes in price level occurs is, because of a time-consuming interest adjustment process and recurring changes in the price level, only a theoretical and not a practically attainable state. Consequently, the modified nominal interest theorem is, in view of economic reality, *incompatible* with the classical postulate of neutrality put forward in monetary theory: even in the longer run, the (expected) real interest rate continues to be influenced in fact, by (expected) price level changes and any monetary changes that lie behind them.

5. Fisher stressed generally the pratical significance of his theorem, combined with a number of detailed statements of the (potentially destabilizing) role of the banking system and of central banks. In particular, he sees a connection between changes in bank reserves (bank balances at the central bank) and inverse changes in nominal market interest rates. Central bank policy can make use of this connection and, via changes in bank

reserves, induce interest rate effects which – according to Fisher and Wicksell – have a stabilizing impact on the price level. Implied here is a modern, interest-orientated concept of money stock control (money "supply" theory). In this connection Fisher also stresses reverse causality: nominal interest rates → price level. Because of the "feedback" relationship described in his theorem (inflation → nominal interest rates), he holds that nominal interest rates cannot *on a long-term view* serve as indicators for central bank policy. From Fisher's observations, we can construct a closed macroeconomic chain of effects running from bank reserves via interest rate effects to changes in the price level, with repercussions on nominal interest rates in accordance with his modified theorem.

## Bibliography

Böhm-Bawerk E., von, *Kapital und Kapitalzins: Positive Theorie des Kapitales*, two volumes (1888), fourth edition, Jena, 1921; reprint, Meisenheim/Glan, 1961.
Bomberger W.A. and Makinen G.E., "The Fisher effect: Graphical treatment and some econometric implications", *Journal of Finance*, June 1977.
Brunner K. and Meltzer A.H., "An aggregative theory for a closed economy", in Stein J.E. (ed.), *Monetarism*, Amsterdam 1976.
Xagan P., "The monetary dynamics of hyperinflation", in Friedman M. (ed.), *Studies in the quantity theory of money*, Chicago 1956.
Carmichael L. und Stebbing P.W., "Fisher's paradox and the theory of interest", *American Economic Review*, September 1983.
Clark J.B., "The gold standard in the light of recent theory", *Political Science Quarterly*, September 1895.
Deutsche Bundesbank, "Die Währungspolitischen, Institutionen und Instrumente in der Bundesrepublik Deutschland", *Sonderdrucke der Deutschen Bundesbank* Nr. 1, Frankfurt, March 1980.
Fisher I., *Appreciation and interest*, New York, 1896; reprint, 1965.
Fisher I., *The nature of capital and income*, New York, 1906.
Fisher I., *The rate of interest*, New York, 1907.
Fisher I., *The purchasing power of money* (1911), second edition, New York, 1922.
Fisher I., "Our unstable dollar and the so-called business cycle", *Journal of the American Statistical Association*, June 1925.
Fisher I., *The money illusion*, New York, 1928.
Fisher I., *The theory of interest*, New York, 1930; reprint, 1970.
Friedman B.M., "Price inflation, portfolio choice, and nominal interest rates", *American Economic Review*, March 1980.
Friedman M., *A theory of the consumption function*, New York, 1957.
Gebauer W., "Geldmarkt", *Handwörterbuch der Wirtschaftswissenschaften*, Volume 3, Stuttgart, 1981.
Gibson W.E., "Price expectations effects on interest rates", *Journal of Finance*, March 1970.
Gibson W.E., "Interest rates and monetary policy", *Journal of Political Economy*, May/June 1970.
Haas J. de, "A third element in the rate of interest, *Journal of the Royal Statistical Society*, March 1889.
Hicks J.R., *Value and capital* (1939), second edition, Oxford 1946.

Hirshleifer J., *Price theory and applications*, Englewood Cliffs (1976), second edition, 1980.
Jevons W.S., *The theory of political economy* (1871), fifth edition, New York, 1957; reprint, 1965.
Jevons W.S., *Investigations in currency and finance*, London, 1884; reprint, New York, 1964.
Keynes J.M., *The general theory of employment, interest and money*, London, 1936.
Keynes J.M., "The theory of the rate of interest", in *The lessons of monetary experience; essays in honour of Irving Fisher*, 1937.
Leijonhufvud A., "The Wicksell connection: Variations on a theme", UCLA, Dept. of Economic Working Paper, Los Angeles, 1979 (reproduced in A. Leijonhufvud, Information and Coordination, New York, Oxford, 1981).
Lerner A.P., *The economics of control*, New York, 1944; reprint, 1971.
Lutz F.A., "The structure of interest rates", *Quarterly Journal of Economics*, November 1940.
Lutz F.A. und Niehans J., "Faktorpreisbindung II: Zinstheorie", in *Handwörterbuch der Wirtschaftswissenschaft*, Volume 2, Stuttgart, 1980.
Modigliani F. und Shiller R.J., "Inflation, rational expectations and the terms structure of interest rates", *Economica*, 1973.
Mundell R.A., "Inflation and real interest", *Journal of Political Economy*, June 1963.
Neumann M.J.M., "Inflationserwartungen und andere Determinanten des Zinsniveaus", *Konjunkturpolitik*, 4. Heft, 1977.
Ozga S.A., *Expectations in economic theory*, London, 1965.
Popper K.R., *Logik der Forschung* (Vienna, 1934), sixth edition, Tübingen, 1976.
Richter R., Schlieper U. und Friedman W., *Makroökonomik* (1973), third edition, Berlin, 1978.
Rist Ch., *History of monetary and credit theory*, New York, 1940; reprint, 1966.
Robertson D.H., "Mr Keynes and the rate of interest", in Robertson D.H., *Essays in monetary theory*, London, 1940.
Rutledge J., *A monetarist model of inflationary expectations*, Lexington, 1974.
Rutledge J., "Irving Fisher and autoregressive expectations", *American Economic Review*, February 1977.
Samuelson P.A., "An exact consumption loan model of interest with or without the social contrievance of money, *Journal of Political Economy*, December 1958.
Sargent Th. J., "Anticipated inflation and the nominal rate of interest", *Quarterly Journal of Economics*, May 1972.
Sargent Th. J., "Interest rates and prices in the long run", *Journal of Money, Credit and Banking*, February 1973.
Sargent Th. J., "Interest rates and expected inflation: A selective summary of recent research", *NBER Occasional Papers*, No 3, 1976.
Sargent Th. J., *Macroeconomic theory*, New York, 1979.
Solow R.M., *Capital theory and the rate of return*, Amsterdam, 1963.
Tobin J., "A general equilibrium approach to monetary theory", *Journal of Money, Credit and Bankig*, February 1969.
Wicksell K., *Vorlesungen über Nationalökonomie*, Volume 2 "Geld und Kredit", Jena, 1922? reprint, Aalen, 1969.
Yohe W.P. und Karnosky D.S., interest rates and price level changes 1952-69, *Federal Reserve Bank of St Louis Review*, December 1969.

## Notes

[1] For an early survey of the main points of the controversy, see Jevons (1884), Wicksell (1922) or Rist (1940).

[2] "To alter the mode of measurement does not alter the actual quantities involved but merely the numbers by which they are represented", Fisher (1896), p. 1.

[3] J. S. Mill (1848), pp. 644-649 ("The rate of interest, how far and in what sense connected with the value of money"), de Hass (1889), Clark (1895).

[4] For further details see Annex to this Chapter.

[5] Fisher (1922).

[6] Or $p^* = (P_1^* - P_0) / P_0$, or $p^* = P_1^*/P_0$.

[7] Generally, $(1+p^*)(1-\pi^*) = 1$, i.e. the relative appreciation of goods and the relative depreciation of money counterbalance one another. Thus, $\pi^* = 1 - (1/(1+p^*)) = p^*/(1+p^*)$.

[8] For a definition of the general institutional arrengements, see for example Deutsche Bundesbank (1980), Chapter I, and, specifically on the money market, Gebauer (1981).

[9] Cf., for example, "forward rates" in the expectation theory of Hicks (1946) and Lutz (1940) concerning the term structure of interest rates.

[10] Ozga (1965), p. 23.

[11] Fisher (1896), Chapter II.

[12] For a detailed analysis of the time aspect, see Annex below.

[13] See Fisher's explanation on his facts (Chapter 2) and its modern extention to an open economy. At Fisher's original assumptions (see above), the "Fisher-open" theorem is identical with the Keynesian theorem of interest-rate parity in arbitrage equilibrium. For an exposition, see Gebauer (1982), pp. 81-87.

[14] Fisher (1930), p. 41.

[15] Fisher (1896), p. 92.

[16] Fisher (1896), p. 30.

[17] Time preference for the present and the productive allocation of resources are basic determinants of interest rates in the modern theory of interest, which essentially goes back to Böhm-Bawerk (1888) and Fisher (1930). These determinants are not sufficient as a condition for non-negative interest rates if the generation aspect and institutional factors are explicitly taken into account; see Samuelson (1958).

[18] The simplified formula has gained widespread currency not only in pratical economic policy, but also in modern academic writing. See, for example, the textbook by Dornbusch and Fisher (1981), Chapters 6 and 13. Exceptions are Richter, Schlieper, Friedman (1978), p. 136.

[19] Where the values of i and $p^*$ are high, the obvious: if, in the above example, the decimal point is moved one figure to the right (inflation rate 50%, nominal interest rate 100% — admittedly a very hypothetical situation), the correct value of the expected real interest rate applying (1.3) is 33% and not 50%.

[20] The correctness of (1.2') and (1.3') in the event of continuous compound interest is easy to demonstrate in formal terms. If the annual rate of interest is i, 1 dollar will grow to $1+i$ dollars over one year. If interest is credited quarterly, 1 dollar will have grown to $(1+(i/4))^4$ dollars by the end of the year, and at any other interest payment interval z to $(1+(i/z))^z$ dollars by the end of the year. If interest crediting is continuous, occurring every second as it were, z will tend towards infinity; therefore:

$$\lim_{z \to \infty} (1+(i/z))^z = \lim_{z' \to \infty} ((1+i/z')^{z'})^i, \quad z' = z/i.$$

Now, $(1+i/z')^{z'}$, with $z' \to \infty$, is equal to the natural number e. One dollar therefore grows to $e^i$ dollars by the end of the year assuming continuous compound interest. Therefore, the theorem (1.1) is now

$$e^{r^*} + e^{p^*} = e^i. \qquad (1.1')$$

By taking the logarithm of (1.1'), we arrive at (1.2') directly. For the derivation, see Fisher (1906 and 1907) and, more recently, Hirshleifer (1980).
[21] For example, by Sargent (1976).
[22] Usually, interest in the case of fairly short-term financial assets (with a maturity of under one year) is calculated and paid upon maturity of the capital sum; in the case of long-term financial assets, interest is generally credited annually.
[23] See above Assumption 1. By contrast, Fisher used the assumption of price stability as the basis for his interest theory proper, which seeks to establish the "fundamental" determinants of the interest rate; see Fisher (1930), p. 46 and p. 494.
[24] I am using here, in slightly adapted form, the chart produced by Fisher (1930), p. 412.
[25] If one were to apply the complete nominal interest theorem (1.2) here, the nominal interest rate would have to become negative in the numerical examle given, a situation which would be inadmissible.
[26] Fisher (1930), pp. 414-415.
[27] Fisher does not give any further explanation of what he means by a "normal" interest rate; however, there is some suggestion of an affinity with Wicksell's "natural" interest rate.
[28] Böhm-Bawerk (1888), Book IV; Fisher (1930), Chapter I.
[29] See Hirshleifer (1980), Chapter 16.
[30] See Richter, Schlieper, Friedman (1978), pp. 135-137. The forward price has the dimension of "dollars payable today per unit of goods available in one year's time".
[31] For discussion of this concept, see, for example, Keynes (1936), p. 222 et seq.
[32] See, for example, the statistics on the lending and deposit rates of credit institutions, regularly published in the Monthly Reports of the Deutsche Bundesbank, Table V.6 (statistical section).
[33] The original work on this point stems from Jevons (1871), pp. 247-250.
[34] "The rate of interest is always relative to the standard in which it is expressed"; Fisher (1930), p. 41.
[35] "Dimension — The kind or species of any magnitude as indicated by its measurement in terms of another magnitude"; Fisher (1906), p. 331.
[36] Keynes (1936), p. 224. If this composite commodity appreciates relative to money by y% annually, the marginal efficiency of capital (defined as x% in units of money) is equal to (x − y)% in units of the composite commodity.
[37] Keynes (1937), p. 145; for a critical examination, see Robertson (1940).
[38] The reciprocal expression $1/J'$ denotes the rate of capitalization in years. The lower the interest rate, therefore, the longer the time-span of capitalization, and vice versa.
[39] Here we can see the connection with the derivation of the Fisher theorem on the basis of one year (see above).
[40] For fuller treatment of this subject, see Fisher (1906), pp. 363-364.
[41] See Fisher's calculations (1906), p. 1908 and p. 362.
[42] "Capital assets" is used here to include generally all physical and non-physical, economically relevant goods. Non-physical capital assets have been introduced into monetary macroeconomics through the concept of "human capital" used by Friedman (1957). The argument, which has been going on for decades, as to whether the term "capital" should not be used instead of "wealth", is not taken into account here, as *analytically* the theory of interest does not necessarily require a theory of capital; see, for example, Solow (1963).
[43] Fisher (1930), p. 15.
[44] See Lerner (1944).
[45] See, for example, the models of Brunner and Meltzer (1976) and of Tobin (1969).
[46] See, for example, the analysis of Leijonhufvud (1979).
[47] "Rent" is the ratio of the payment to the physical object, e.g. dollars per acre of land. "Interest" in the sense of interest yield is the ratio of payment to the *value* of physical

objects, e.g. dollars per present value (in dollars) of a house.

[48] Hirshleifer (1980), Chapter 15.

[49] For example, in Hirshleifer (1980), pp. 500-501 or in Lutz and Niehans (1980), pp. 530-531.

[50] Fisher (1930), p. 399.

[51] Fisher (1896), pp. 36-37.

[52] Fisher (1930), p. 400.

[53] After the introductory remarks quoted from Chapter XIX in his *Theory of interest* (1930, Fisher continues: "The main object of this chapter is to ascertain to what extent, if at all, a change in the general price level actually affects the market rates of interest" (p. 399). There is no mention of expectations. Again, by way of introduction to his correlation analysis, he states that "... the theory being investigated is that interest rates move in the opposite direction to changes in the value of money, that is, in the same direction as price changes ..." (p. 416).

[54] Recent exceptions are Rutledge (1977) and B. Friedman (March 1980); see also Gebauer (1976) for a critique of the usual interpretation of Fisher.

[55] Fisher (1896), pp. 38-53; Fisher (1930), pp. 401-407.

[56] The interest payments and redemptions in the case of the rupee bonds were made in London in sterling; the amount payable in sterling therefore depended on the exchange-rate relationship between sterling and the rupee.

[57] Fisher, 1930, pp. 401-402. Obviously, Fisher is assuming here that there is in the eyes of market participants a "normal" or "appropriate" value ratio between the two currency units. Such an assumption is characterized today by the term "regressive expectation formation". For a more specific or formal description of this idea see Keynes' liquidity preference theory (1936) and the "preferred habit" theory of yield curve analysis by modigliani and Shiller (1973).

[58] In fact, the price of gold fell from almost 120 dollars per unit of gold in 1870 to parity at 100 dollars in 1879, when the Greenback period ended and specie payments in gold were officially resumed. Gold did in fact, as obviously expected by the market, loose value relative to paper currency.

[59] See Fama (1970) and Fisher (1980).

[60] As already stated, the level of the interest and redemption payments actually payable depended on this exchange-rate relationship.

[61] Fisher (1930), pp. 405-406.

[62] Fisher (1930), p. 405.

[63] In his introductory remarks to these comparisons, the subject of investigation was indicated as follows: "Evidence that an expected change in the price level does have an effect on the money rate of interest may be obtained from several sources ..."; Fisher (1930), p. 400.

[64] Fisher (1896), p. 75.

[65] Fisher (1930), p. 407.

[66] Fisher (1896), pp. 54-57; Fisher (1930), pp. 407-416. Later calculations in Fisher's *Theory of interest* (1930) combine, with some additions and amendments, the initial factual evidence set out in *Appreciation and interest*. I therefore refer below essentially to the relevant passages in Fisher's *Theory of interest* (1930).

[67] Fisher (1930), pp. 407-408 (underlining by Fisher).

[68] Fisher (1940), p. 526.

[69] London, New York, Berlin, Paris, Calcutta and Tokyo.

[70] The longest time series (yearly data for Great Britain) spans a full century, and the shortest time series (yearly data for Japan) 40 Years. See Fisher (1930), Appendix, pp. 520-529.

[71] More suitable price indices such as the implicit price deflator for GNP or the cost of living index for all private households were not available to Fisher. More recent calculations, for example, for the Federal Republic of Germany, mostly use such indices; see Gebauer (1976).

[72] For example, the total observation period for British data (london capital market) covers the years 1824-1927; the time series is then broken down into successive sub-periods of varying length, beginning with a 10-year period 1825-34, followed by a five-year period 1834-39, then a 12-year period 1839-52, etc.; Fisher (1930), Appendix, p. 527.

[73] Fisher (1930), p. 527, Table VII.

[74] Fisher (1930), p. 408; on the macroeconomic interpretation of the Fisher theorem, see Chapter 3 below.

[75] Fisher (1930), p. 410.

[76] In the sense of Popper (1976). See also Carmichael and Stebbing (1983).

[77] "The evidence obtained ... indicates that there is a very apparent, though feeble, tendency for the interest rate to be high when prices are rising, and the reverse"; Fisher (1930), p. 411.

[78] Fisher (1930), pp. 411 and 416 (underlining by Fisher).

[79] For example, standard deviations of real interest rates are between seven and thirteen times higher than the standard deviations of the corresponding nominal interest rates; Fisher (1930), p. 415, Table 14.

[80] Fisher (1930), Appendix p. 527, Table VII.

[81] Fisher (1930), p. 413.

[82] Fisher (1930), pp. 415 and 494.

[83] Fisher (1930), pp. 399-400 and 415.

[84] Fisher (1928). For a discussion of the concept, see, for example, Badura (1977), p. 88 et seq.

[85] Which is, interestingly enough, the definition of the "classical dichotomy" or "neutrality of money". For more recent definitions of the "money illusion", see, for example, Dornbusch and Fisher (1981), p. 214, or Richter, Schlieper and Friedman (1978), p. 124.

[86] On further macroeconomic analysis and application, see, for example, Sargent (1979).

[87] Fisher (1896), pp. 76-79 and Fisher (1922), p. 55 et seq.

[88] Fisher (1930), p. 494.

[89] Fisher (1930), p. 416.

[90] The British yearly data span more than a century, i.e. from 1820 to 1924, and the US yearly data run from 1900 to 1927; see Fisher (1930), Appendix, pp. 530-531.

[91] Fisher (1930), p. 418, Chart 45.

[92] Fisher (1930), p. 418; the correlation coefficients in Fisher's Chart 45 (p. 418) are very low and are in some cases nearly zero.

[93] Division by the sum of the weights, which is normally necessary, is not carried out because of the specification $\Sigma w_l = 1$. A numerical example may clarify Fisher's use of (2.1): with an assumed ditributed lag of $n = 8$ years from $l = 1$ to $l = 9$ and arithmatically decreasing weights $w_1 = 8/36$, $w_2 = 7/36$, ..., $w_8 = 1/36$, $w_9 = 0$, we have

$$\sum_{l=1}^{9} w_l = 1;$$

the weighted average $\bar{p}$ is obtained from summing

$$(8/36)\, p_{-1} + (7/36)\, p_{-2} + \ldots + (1/36)\, p_{-8} = \sum_{l=1}^{9} w_l p_{t-l} = \bar{p}.$$

[94] For a broader macroecomic interpretation of Fisher, see Chapter 3 below.

[95] For example, Sargent (February 1973), p. 387, tests a regression formula and explains misleadingly: "... which is the equation that Fisher implemented in his empirical work".

[96] As put forward, for example, by Neumann (1977).

[97] The connection between regression and correlation analysis may be used only to translate Fisher's correlation coefficient into the coefficient of determination of a — hypothetical — regression formula, and Fisher's method of handling the weights in (2.1) *may* be enlisted as the (necessary) identifying restriction for a parameter estimated by regression analysis; see Yohe and Karnosky (1969), p. 20

[98] Early examples are the interpretations by Hamburger and Silber (1969), Yohe and Karnosky (1969), Gibson (March 1970 and June 1970), Feldstein and Eckstein (1970), Sargent (1972 and February 1973), who a decade ago gave expression to the renewed interest in Fisher's theorem and the distributed lag idea in connection with the debate on modern monetarism.

[99] Cagan (1956) and Friedman (1957). A typical extract: "Anticipating the work ... of Cagan and Friedman by about 25 years, Fisher posited that people form expectations by taking a weighted sum of current and past actual rates of inflation ..."; Sargent (February 1973), p. 386. Moreover, Sargent's reference to "current rates of inflation" is incorrect; Fisher ignored the current period; in Formula (2.1) the time index runs from 1 to n, and not from 0 to n.

[100] Such exceptions are, for example, Mundell (1963) and, more recently, Dornbusch and Fisher (1981).

[101] The time index t denotes the current period; it was superfluous in the original theorem, which describes a static situation of equilibrium, ad was omitted by Fisher (and likewise in the above presentation in Section 1.2).

[102] A numerical example: let us assume that $i_t = 10\%$, $r^* = 5\%$ and $p_t^* = 5\%$. Let us also assume that in the next period $t+1$ inflationary expectations of 5% rise to 7%. If the nominal interest rate is now adjusted initially only imperfectly because of (2.3), for example by $\Delta i = 1\%$ point, the real interest rate must fall by $-\Delta r^* = 1\%$ point, so as to comply with equation (1.2'). In period $t+1$, the following equation then applies:

$$i_{t+1}\,(11\%) = r_{t+1}^*\,(4\%) + p_{t+1}^*\,(7\%).$$

[103] An interpretation emphasizing the time horizon may be found in Rutledge (1974 and 1977), Bomberger and Makinen (1977) and Dornbusch and Fisher (1981).

[104] Mundell (1963 and 1971, Chapter 2).

[105] Keynes (1936), p. 142.

[106] The mean lag is simply the weighted average lag; it shows the time which elapses until half of the influence of the independent variables (in this case, past inflation rates) on the dependent variable (in this case, nominal interest rate) has taken place. Since in Fisher's calculation method the weights add to one, he was able to calculate the mean lag L in accordance with the simple formula $L = (n-1)/3$ or approximately $L = n/3$. For a basic definition of the mean lag, see Griliches (1967), p. 31.

[107] Fisher (1930), Charts 46 and 47, pp. 421-422.

[108] Fisher (1930), pp. 423-425.

[109] Fisher (1930), p. 425.

[110] Fisher (1930), pp. 425-429 and Appendix, pp. 532-533.

[111] Fisher (1930), p. 428 (underlining by Fisher).

[112] Gibson (1923).

[113] Keynes (1930, pp. 198-210. The Gibson phenomenon was in the last facts, but not with the theoretical "classical dichotomy".

[114] Fisher (1930), p. 430.
[115] Fisher (1930), pp. 431-438.
[116] Fisher (1930), p. 438.
[117] Fisher (1930), p. 440.
[118] Fisher (1911). The distinction between a "short-term" transition period and a "long-run" state of equilibrium is of crucial importance in Fisher's reformulation of the quantity theory of money: this theory and the famous quantity equation associated with it apply *only* in a state of long-term equilibrium.
[119] Fisher (1930), p. 441 (my underlining).
[120] Fisher (1930), pp. 440-441. As an example, Fisher refers to the "absurd" idea that nominal interest could rise to a higher level "if ... we were to call a cent a dollar asnd thereby raise the price level a hundredfold" (op. cit.).
[121] See, for example, Meiselmann (1963), Friedman and Schwartz (1976) and Shiller and Siegel (1977) for the United States, Badura (1977) and Sauer (1977) for the Federal Republic of Germany and Fase (1975) for the netherlands. The list could be continued for other countries.
[122] For example, by Shiller and Siegel (1977).
[123] I am thinking here of Shiller and Siegel (1977), p. 896 et seq., and Sargent (February 1973), pp. 386-387. A typical statement is: "However, it is possible to argue that Fisher's explanation of the Gibson paradox is really only a redefinition of it" (Sargent, February 1973, p. 387).
[124] Fisher (1930), p. 428.
[125] Fisher (1930), p. 429 et seq., in particular pp. 439-440.
[126] Assuming a given amount of reserves in the banking system; see Fisher (1896), pp. 76-79. The argument is obviously along the lines of Wicksell (1898) and (later) Keynes (1936).
[127] Fisher (1930), p. 439.
[128] Fisher (1922).
[129] See in particular pp. 56-60.
[130] Fisher (1922), p. 60. For a recent presentation and analysis of Fisher's monetary business cycle theory, see Bomberger and Makinen (1977); on the interaction of nominal and real interest rate effects, see Carr and Smith (1972).
[131] Fisher (1922), p. 60 (underlining in the original).
[132] Friedman (March 1980), p. 32; see also Rutledge (1977).
[133] It should be noted that in 1926, i.e. long before A. W. Phillips (1958), Fisher postulated and provided statistical evidence of the relationship which has been named after Phillips. Fisher postulated the relationship as a *short-term* disequilibrium phenomenon. It would therefore be more correct to call the short-term relationship between inflation and unemployment the *Fisher-Phillips curve*.
[134] Fisher (1926), p. 786; reprinted in the Journal of Political Economy (1973), p. 497 et seq. (quotation from p. 498). Fisher continues: "The *ultimate* effects of a long-continued inflation are doubless bad all round ..." (my underlining).
[135] Friedman (1969); for a survey, see Gordon (1976 and 1978), Frisch (1980) and Dornbusch and Fisher (1981). One of Friedman's key sentences, in concurrence with Fisher, is as follows: "... there is always a temporary trade-off between inflation and unemployment; there is no permanent trade-off"; Friedman (1969), p. 104.
[136] Fisher (1925) and Fisher (1926).
[137] Fisher (1930), p. 440.
[138] Fisher (1930), p. 442.
[139] Fisher (1930), p. 443 and footnote 21.
[140] "In fact, an arbitrary increase in i at any time does tend to pull down the level of general commodity prices, while a decrease in i tends to increase P"; Fisher (1930), p. 443; see

also the passages from Fisger's *Purchasing power of money* mentioned in the previous section.

[141] Fisher (1930), p. 443.

[142] See Ehrlicher and Oberhauser (1978) and the further bibliographical references given there.

[143] Monetaru control (1980).

[144] On the controversy surrouding money stock control, see Federal Reserve Bank of Boston (1980) for an American point of view.

[145] Fisher (1930), p. 444.

[146] Friedman (1969).

[147] Fisher (1930), pp. 444-450.

[148] In a German context, bank reserves can be taken here to mean "credit institutions' balances at the central bank" (and *not* the central bank money stock target at constant reserve ratios).

[149] "Banking thus becomes... a most powerful independent influence"; Fisher (1930), p. 448.

[150] See, for example, Fisher (1925).

[151] If one regards the central bank as the control institution with responsability for the banking system; see Friedman (1959).

[152] The complete chain of causation is then: bank reserves → nominal interest rates → money stock → business cycle → prices, price expectations [→ credit demand → bank reserves] → nominal interest rates ...

[153] Fisher (1930), pp. 46-47 and p. 447. Without specifying or indeed taking full account of Fisher's original contribution, the question of the "price of money" has recently been taken up by monetarists and introduced into the debate; see the argument between Friedman and Tobin in Stein (1976). Generally speaking, Fisher felt that such oversimplifications were misleading rather than englightening..

[154] The reader is referred to the Annex to Chapter 1 above.

[155] See Annex to Chapter 1 above.

[156] "... changes in the purchasing power of money tend ... to affect the nominal rate of interest in one direction and the real rate of interest in the opposite direction"; Fisher (1930), p. 505.

[157] Fisher (1930), p. 494.

[158] "The business man supposes he makes his contracts in a certain rate of interest, only to wake up later and find that, in terms of real goods, the rate is quite different"; Fisher (1930), p. 44.

[159] Fisher (1930), pp. 43-44.

# Economic Crisis in Eastern Europe: Prospects and Repercussions

DOMENICO MARIO NUTI
Birmingham/Florence

## 1. Introduction

There is an old Polish saying that the main advantage of socialist planning is that it solves successfully all those problems which without socialist planning would not arise. This may appear as an instance of Polish cynicism but in reality is both an understatement and an overstatement. It is an understatement because it neglects the contribution of socialist planning to the industrialization, urbanization, price stability, full employment and social progress of one third of the globe; since the last World War there has been no period in which the growth of income and industrial output in the socialist countries (or "centrally-planned economies" in United Nations parlance) as a group has been exceeded by the groups of advanced capitalist economies or developing countries. It is also, however, an overstatement because the development process of socialist countries has generated new problems, conflicts and contradictions which the system has not yet been able to solve. Since the mid-1970s a crisis has developed in Eastern Europe and the Soviet Union, not only in the broad sense of deterioration of economic performance, stagnation or even decline, but in the stricter sense of a phase of a cycle, due to the relative weight of demands and resources, and deeply rooted in the specific systemic features of these economies, namely the centralization of political power and economic decision-making, and the systemic commitment to economic growth, full employment and price stability.

In my lecture I review the symptoms of this crisis, outline an explanatory model of the functioning of these economies and how they generate crises; I then consider the implications of current trends for the West in general and Western Europe in particular.

## 2. Symptoms of economic crisis

In its most spectacular form economic crisis has manifested itself in Poland: accelerating decline of income, down by 15% in 1981 and still

falling, 40% unutilized capacity and rising redundancy of labour, mounting external indebtedness of the order of $ 30 billion with over 100% debt-service ratio and *de facto* default in international financial commitments, open and repressed inflation at accelerating rates, endemic and growing shortages, rationing and queues, the collapse of the central management of the economy. The economic crisis gave rise to massive political unrest, the conquest of political rights unprecedented in Eastern Europe, and then military rule. Let me say now that Poland is a unique phenomenon deeply rooted in a thousand years of national history and not to be repeated in the foreseeable future in the rest of the socialist commonwealth. But the same kind of symptoms, to a lesser and varied extent, have manifested themselves throughout the area, and can be grouped under the headings of: (i) deteriorating growth performance, (ii) external imbalance and (iii) internal imbalance; with equally important political implications, though less drastic than in Poland.

(i) The deteriorating growth performance is clear from yearly plans in 1976-81 envisaging income growth rates lower than originally postulated in the five-year plans, and at a decreasing rate over the period; and from actual growth rates being systematically lower than planned and also recording a decreasing trend over the period. The only exception is 1977 in Hungary and the German Democratic Republic, where plans slightly more ambitious than in the previous year were realized. It is true that Bulgaria and the GDR maintained their growth, in spite of plan underfulfilment, around the appreciable rates of 6 and 4% a year; but the slowdown was much steeper in other countries (in 1981 growth rates were 3.1% in the Soviet Union, 2.1% in Romania, 1.8% in Hungary, 0.2% in Czechoslovakia) and negative growth was recorded not only in Poland in 1979-81 but also in Hungary in 1980 (-0.8%), far from the earlier record, five-year plans and scaled-down yearly targets. Identical trends can be found in industrial production; while agricultural output, which in the 10 years up to 1983 had increased at an average 5% per year, in the following five years slowed down to 3%, declined in 1979, to recover slightly in 1980-81. Current plans confirm these trends.

This deteriorating growth performance is particularly disappointing in view of the massive accumulation of capital undertaken throughout the area, with the share of gross capital formation in national income rising until the mid-1970s up to levels of 30-40%, with peaks of 45% (Poland 1975), and the policy of technology transfer from the West pursued by means of massive imports of machinery by the Soviet Union and other countries (especially Poland and Hungary, though much less so by the more advanded GDR and Czechoslovakia).

The mirror image of this performance is the generalized increase in capital/output ratios, the slowdown and occasional fall in labour productivity, the increasing inefficiency in the allocation of resources. Microecono-

mic inefficiency has always been a widely recognized feature of the socialist economy, thought to be compensated for by macroeconomic rationality, but the suggestion of increasing overall inefficiency is substantiated by the observation of increasing overcapacity, i.e. unemployment of capital, the dispersion of investment means over a widening front of unfinished projects, and unused or even uninstalled machinery (subject to obsolescence and physical deterioration) often imported at great expense (Poland, Czechoslovakia). There are other concrete indications of increasing inefficiency, such as the continued growth in energy consumption per capita, at 3.2% per year in 1974-78 (with respect to 3.6% in 1971-73), whereas in the same period it fell by 0.2% in developed market economies (with respect to a 3.1% yearly increase in 1971-73); leading to a consumption of energy per unit of output in CMEA countries roughly double that of OECD countries by the end of the decade; clearly the failure to respond to increasing energy scarcity can only be interpreted as increasing inefficiency.

(ii) The second symptom of economic crisis is the widening external imbalance, in the form of recurring trade deficits and mounting cumulative foreign indebtedness. In 1976-81 the trade balance of the six Eastern European countries with non-CMEA countries has been negative every year (with the exception of Bulgaria in 1977-80 and Czechoslovakia in 1980). In 1970-79 the Soviet Union has run trade surpluses with developing countries and slightly smaller trade deficits with the developed West, but since only about a third of its surplus was in convertible currencies it has needed hard currency finance. In 1980 the Soviet Union registered over $ 200 million surplus with the developed West (and $ 2.7 billion with less developed countries) but in 1981 the situation changed drastically, leading to a $ 3 billion deficit with the West and a $ 0.4 billion deficit with less developed countries. Between 1976 and 1981 the total hard currency debt of the six Eastern European currencies more than trebled, to $ 65 billion (about 10 times their total debt in the 1970s; the prudent limit of 40% debt-service ratio recommended by the IMF was overstepped in 1980 by Poland, Hungary, the GDR and Bulgaria, while Poland in 1981-82 and Romania in 1982 have effectively defaulted, though have not been declared officially to have done so. Together with the Soviet Union, also facing a severe cash shortage, total debt exceeds $ 75 billion, while access to fresh credits and even rolling-over facilities is limited by bankers' reluctance to raise their exposure, due to political pressure and loss of confidence in the so-called "umbrella theory", whereby the USSR would ultimately guarantee loans to the entire area. The burden of debt is exacerbated by its time-structure dominated by short-term maturities which have to be renewed at rising interest rates and subject these countries to the continuous scrutiny of their affairs by bankers; and by the combination of inflation and high interest rates which involves "front-

loading" or the "bunching" of the service burden in the early years of indebtedness, with respect to the burden which the same real interest rate would have for a lower rate of inflation.

While it is inappropriate to single out individual trade flows in order to explain the rising indebtedness, it is interesting to note the gradually falling energy net export (mostly Soviet), the mounting food deficit (from 1976 onwards in the six East European countries, from 1972 onwards in the area including the Soviet Union), the large-scale machinery imports by CMEA countries (quintupled over the 1970s for a cumulative amount of almost $ 100 billion, though slightly declining since 1978), the increasing East European dependence on the West for raw materials and semi-finished products (as well as massive hidden subsidies by the Soviet Union to East European countries through the underpricing of Soviet oil exports to them caused by CMEA pricing rules for infrabloc trade).

(iii) Internal imbalance has taken several alternative or simultaneous forms: open inflation; hidden inflation due to forced substitution of cheap products, rationed or simply unavailable, by more expensive substitutes, whether provided by the official or the "unofficial" or "black" markets; repressed inflation, in the form of large and increasing excess liquid balances in the hands of the population, as well as a progressive squeeze on the share of gross profits on turnover. Since the post-war monetary reforms, in 1955-75 open inflation had become an almost forgotten phenomenon in Eastern Europe and the Soviet Union, but in the second half of the 1970s (more recently in the Soviet Union and the GDR) it has returned at a slow but increasing pace (especially if hidden price rises are included); though hyperinflationary rates have only been seen in Poland after the military attempted the restoration of monetary equilibrium. The question "Will there be money under communism?" has two alternative answers in Russian folklore: "There will only be money" and "There will be money for some but not for others"; in the last few years socialist countries have come perilously close to both answers, with endemic shortages of essential goods as well as luxuries, and unequal access to scarce goods through either official privilege or unofficial exchange.

In each period the current imbalance between intended expenditure and the mass of goods and services provided for the population, valued at official prices – the "inflationary gap" – is added to the "inflationary overhang" that is the sum of past inflationary gaps cumulated over time. While the precise calculation of intended expenditure (and therefore of inflationary gaps) raises controversial methodological problems, and the mere presence of excess demand (queues, rationing) should not necessarily be identified with repressed inflation (which is an *increase* in the rate of excess demand) we can nevertheless, from the observed rise in monetary balances held by the population with respect to current income, infer a progressively widening gap between actual and desired monetary balances

of the population. Even for a falling inflationary gap the inflationary overhang increases over time, and this cumulative problem exercises a destabilizing influence on the orderly distribution of consumption goods and services; shortages have a strong demoralizing effect and adversely affect productivity through time-wasting search and queueing and through the weakening of monetary incentives; moreover, they lead to the population hoarding more consumption goods and distort consumption patterns. Sooner or later, through derived demand for competing inputs or through international trade, inflationary pressure spills over from the consumption goods sector to production goods, leading to shortages of supplies to production enterprises. Shortages have a multiplier effect, in that one rouble of missing component may lead to a chain reaction of disruptions in other supplies to production enterprises down the line; this "supply multiplier" effect is growing all the time with the growing complexity of internal and international division of labour, and is only alleviated by a tendency towards vertical integration within the industrial system (which is most advanced in the GDR model but is an observable trend throughout the area). Shortages also promote the growth of the non-socialist sector, composed of a private sector, a straight criminal economic sector, as well as a "second" or "informal" sector in different shades of grey, intermediate between the official private and the criminal sphere, where unplanned (or rather a-planned) transactions take place. This "second" economy, however, is an integral part of the system, which is why it is tolerated and occasionally encouraged. There are innumerable examples of the "privatization" of State power on the part of people invested with that power, for their own benefit and that of their protégés. The problem is not the excessive power of the State, as claimed by critics of Soviet-type alleged "State capitalism" or "State socialism", but a quasi-feudal fragmentation and devolution of State power, with interlocking small and large fiefs rigidly set in an ossified structure, at odds with social justice and economic efficiency.

The economic crisis and its visible side-effects generate, at the same time, pressure for economic and political change. For a quarter of a century there have been numerous attempts, in the Eastern European countries and the Soviet Union, at relaxing central planning and introducing greater enterprise autonomy; market rather than planned discipline; profit-linked investment and material incentives. The extraordinary thing is that these attempts have not succeeded and reform projects either remained dead letter or failed and were reversed. (The only exception seems to be Hungary, where the so-called New Economic Mechanism was introduced in 1968 and has done relatively well, though the system, even there, has a tendency to revert to the original Soviet-type model under economic pressure.) The same thing is true of political reform: economic failure undermines the legitimacy of the system, breeds dissent and leads

to pressure for political democratization and liberalization; historically, moves in this direction accompany, with leads or lags, trends towards economic reform, and also revert towards the authoritarian model sooner or later. Poland, again, is the most spectacular instance of such a reversal.

We now have a number of important questions: Why is there an economic crisis in Eastern Europe? Why is there no stable progress towards economic and political reform? Can the system generate alternative solutions? Or is it doomed?

## 3. A model of socialist development

Soviet and East European literature talks of "disproportions" and "imbalances" instead of crisis, and relies primarily on three explanations : (i) exogenous factors, such as natural disasters and the world crisis; (ii) the wickedness or incompetence of individuals, to be found and punished; (iii) the switch from intensive to intensive growth in East European development. There is no doubt that "crisis" is what these dubious euphemisms stand for, and that the factors suggested, while having substantial weight, do not add up to a total explanation.

Natural disasters like droughts and floods, ice and heat waves beyond reasonable expectations have taken place and have had an adverse effect (especially in Poland and the USSR) on harvests, building and transport, but they were not so unpredictable; moreover, their effects have been amplified by persistent underinvestment in these sectors and by systemic problems, especially in agriculture. The energy crisis has affected the export prospects of East European countries but they have shown no flexibility in adapting to new conditions; besides, the Soviet Union is a net oil exporter and directly benefited through its terms of trade, Poland was a net energy exporter and through coal sustained its terms of trade (actually improving them in 1974), and East European countries had at least three years' notice before being exposed to dearer oil. The incompetence of decision-makers must be regarded as a systemic feature in view of the numbers involved, and a selection system based on party nomination to crucial posts (*nomenklatura*) instead of professional merit. As for the switch to intensive development, the exhaustion of labour reserves and natural resources has certainly raised the cost of growth and possibly acted as a triggering factor for economic problems: however, if this were all it would be difficult to understand why economic reforms have failed to produce a system more suitable to the "intensive" stage of socialist development.

I suggest a more complex model of socialist development which embodies these factors, and others, in a pattern of behavioural regularities, conflicts and contradictions, which are capable of generating a crisis of the

type just reviewed, as well as a more attractive potential course. It is a model of the possible "laws of motion" of a socialist system, which applies, among other things, Marxian methods to a Marxian-inspired system, with a vengeance.

The construction of the model proceedes from the following empirical observations, or "stylized facts" of socialist development:

(i) There is a connection between political centralization and economic centralization; to start with the establishment of one-party rule and "democratic centralism" leads to the centralized Soviet-type model; subsequently forms of political democratization are "associated with" (i.e. without asserting a one-directional causal link) economic decentralization.

(ii) Economic centralization is associated with microeconomic inefficiency in the allocation of resources, for the choice of production methods, their technical application, and the choice of consumption structure.

(iii) Economic and political centralization leads to a bias towards capital accumulation. First, at all levels the same urge to accumulate typical of the capitalist system is present, but without the checks and constraints of that system (such as stock market valuation of enterprise assets, takeover bids, bankruptcy discipline). Second, overoptimistic expectations about the system's ability to promote labour productivity raise investment requirements for the creation of new jobs over actual requirements. Third, in the course of communications and interactions between the various agents (Party, Gosplan, Ministries, firms associations and firms) strategies are developed that result in ambitious investment policies.

(iv) As long as there are labour and material reserves, the capital accumulation bias of the system generates fast economic growth, outweighing the disadvantages of microeconomic inefficiency and ultimately leading to "full employment", in the strict sense of exhaustion of labour and material reserves (thus, for instance, not only full employment of those willing and able to work, but high industrialization and urbanization).

(v) At "full employment", so defined, capital accumulation comes up against increasing costs of economic growth; also when the full employment is hit there are echo repercussions throughout the economy which make for fluctuations and some instability. The microeconomic inefficiency of the system becomes more serious, in that at full employment wasting the same amount of labour and materials becomes "more" inefficient than before full employment. Thus full employment adds to pressure for economic reform. In principle, an economic reform could take place, encouraging political democratization which in turn would encourage further decentralization and so on until market socialism and socialist democracy are established, in the best of all possible worlds. Why does this not happen? Because the accumulation bias and full employment also have adverse effects.

(vi) In the absence of automatic checks, the persistence of accumulation bias at full employment causes overaccumulation. Accumulation can be excessive not, or not only, with respect to the preferences of the population, or other plausible but intangible benchmarks; accumulation can be said to have been excessive in Eastern Europe and the Soviet Union with respect to the interests of maintainable consumption (investment leading to higher income streams but not higher consumption streams in the future); with respect to the absorption capacity of the system (investment being "frozen" in work-in-progress for longer gestation periods than necessary, or machines uninstalled or unused); with respect to the ability of the system of financing both internally and externally the capital expenditure and the recurring costs of operation; with respect to the preferences expressed by the leadership in their own plans, since often the sum of investment plans adds up to more than the centrally planned level of investment, and those plans are systematically overfulfilled. Even when investment is slowed down (as in the second half of the 1970s), unless the slowdown is planned long in advance (which it was not), investment continues to be excessive and contributes to excess demand.

(vii) Full employment of labour, in spite of wage norms and an incomes policy directed at maintaining broad macroeconomic balance in consumption markets, leads to wage drift (through informal bargaining at the factory level) and adds to excess demand.

(viii) The built-in tendencies towards excess demand in both production and consumption markets are aggravated, or alleviated, by the net effects of exogenous factors and their change over time: the volume of world trade, terms of trade, the volume and terms of internatinal loans, defence expenditure, agricultural harvests, the pace of technical advance, the exhaustion or discovery of natural resources. It would require a permanently favourable balance of these factors to offset the underlying excess demand, which is utterly unrealistic.

(ix) In view of the systemic commitment to price stability, enshrined in the theory and practice of socialist countries where prices are raised only under the most formidable demand and cost pressure, normally excess demand generates shortages, which cumulate over time and through spillover effects disrupt the system of management and planning of the national economy; shortages and disruption generate pressure towards greater and not lesser centralization, and the application of war-type methods of rationing, priorities and direct control.

(x) The coexistence of pressure to reform (due to inefficiency and instability), shortages and pressure to centralize leads to an impasse: either the pressure to centralize is stronger than the reform drive, and there is no reform, or the reform drive is stronger and a reform takes place in conditions of shortages and excess demand which doom it to fail. The system is locked into a vicious circle of stagnation and waste.

This is not, however, the end of the story. Shortages, inefficiency and instability lead to mounting discontent and political unrest. This eventually affects the level of political centralization, but the impact can go either way: sustained political pressure may induce democratization, but if pressure goes beyond a critical point it is perceived as a challenge to the system and it may well induce an authoritarian involution. The lesson of the recent Polish experience is not that political change is impossible, but that it is touch and go whether sustained political pressure leads to democratization or military rule.

The system can, or at any rate it would be rash and unwarranted to claim that it cannot, break out of its vicious circle by the weakening of any of the perverse links: autonomous political democratization as in the Soviet Union in 1956 or Czechoslovakia in 1967-68; economic reforms timed to coincide with favourable exogenous factors, as in Hungary in 1968; a relaxation of the system's commitment to full employment, as advocated by Dr Popov of the Moscow Academy of Sciences Institute of Management; a relaxation of the system's commitment to price stability, which seems to be taking place throughout the area, and which would prepare a better environment for economic reform; last but not least the possibility of political pressure being directed not at opposing the system but at negotiating political democratization in exchange for the acceptance of austerity, as might have happened in Poland if both trade unionists and government had had more sense. It may be overoptimistic to expect rapid and drastic changes along these lines in the near future in Eastern Europe and the Soviet Union, but it is plainly wrong the expect that the system is incapable of changing, or to expect its collapse.

## 4. Implications for the West

The same kind of symptoms of economic crisis and prospects are also shared by and large by capitalist countries: deterioration of growth performance, external and internal imbalance; it is also a cyclical phenomenon rooted in the systemic features of capitalism. Some of the causes are common to both groups: the world energy crisis and concomitant inflation and recession in world trade; poor labour relations and the failure to reach a social pact between government and workers. But other causes, the actual mechanisms, depth and specific features of the crises, however, are profoundly different. In particular, capitalist countries face macroeconomic constraints internally and internationally instead of structural capacity constraints; they have unemployment of labour as well as unemployment of capital, and instead of several queues for commodities they have one single and long queue for jobs, people being unable to buy money with their labour (which is even worse than being unable to buy goods with money because unspent labour cannot be saved for later expenditure).

These specific differences suggest that there is a certain degree of complementarity between the crises affecting different systems. This complementarity generates opportunities for foreign trade and capital movements of a kind which has nothing to do with conventional Heckscher-Ohlin type international trade theory, i.e. with relative factor endowments and even relative market prices, but is nevertheless a tangible form of comparative advantages. The sale by the West of advanced technology, grain, foodstuffs, steel, semifinished goods, contributes to solving East European structural problems of mismatching of demand and supply, while the sale of *anything* can relieve the overall pressure of excess demand discussed earlier, at a lower social than private cost to the West in view of large-scale unemployment and spare capacity. These sales can activate capacity utilization in the East, with supply-multiplier positive effects, and generate additional goods which would enable the East to recover and pay, sooner or later, for what they have received. The problem is that short of mounting a gigantic multilateral intertemporal barter operation, which is impracticable, the benefits of this complementarity between economic systems can only be reaped by the provision of large-scale credit, which is not impossible but requires a degree of international cooperation, goodwill, and development of financial institutions which is simply not there. Economic factors, however, would amply reward East and West for any effort aimed at trade promotion and credit expansion.

Any move in the opposite direction, in the form of trade embargoes, credit restrictions, refusal to reschedule debt and demands for formal declarations of default for the worst affected countries like Poland and Romania, has an economic cost for both the West and the East and has a destabilizing effect of the economic/political cycle of socialist countries, through the aggravation of shortages, inefficiencies and economic disorganization. President Reagan believes that the economic cost to the West of trade denial and credit denial in economic relations with the East is worth paying in order to expose and enhance the economic difficulties of the socialist bloc, to force political democratization in socialist countries, especially the USSR and Poland. On this, three final observations are offered:

First, the cause of democratization in Eastern Europe is better served by trade and credit expansion than restriction; the economic difficulties of these countries are already sufficient to generate political pressure. By aggravating these difficulties through arms races and economic warfare the diffusion of authoritarian solutions throughout the bloc is more likely than democratization. If, for instance, Western assistance had been made available to Poland in 1981, instead of Western banks cutting almost their entire short-term credits in the spring of that year, the Polish economy would not have collapsed and the state of emergency would have been appreciably less likely.

Second, while both East and West stand to gain considerably from economic cooperation in trade and loans, Western economic warfare could cost the West a great deal and would not hurt socialist economies much. For instance, embargo on the Soviet gas pipeline construction would cost the West at least $ 10 billion machinery exports but would only temporarily delay the project, actually improving Soviet cash flow in the short run. High technology sales to the Soviet Union no longer contribute as much as they used to to Soviet productivity growth. The formal declaration of Polish default would seriously affect Western banking circles (especially if the possibility of deposit withdrawals from the most affected banks are considered) and lead to the write-off of Polish debt and the loss of current interest payments, but Poland could continue to trade on a balanced basis for cash, through friendly intermediaries and alternative transport routes, the higher cost of intermediation being offset by the relief from interest payments. In spite of appearances, the socialist countries are not so vulnerable to Western warfare.

Finally, the considerable cost to the West of trade and credit denial would fall much more heavily on Western Europe than on the United States. For instance, out of the $ 25 billion of Polish debt to the West, less than $ 4 billion represent the maximum US exposure both direct and through government guarantees. The bulk of machinery sales for the Soviet gas pipeline would be commissioned to Europe, while Soviet gas would reduce European dependence from OPEC countries and broaden European energy options. US exports of high technology to the USSR are of the order of $ 150 million a year, i.e. half those of Italy and France, and a fifth of West German sales. The disparity between the USA and Western Europe is just as bad in the general field of manufacturing products. There is a significant unfairness in the distribution of costs of Reaganite policies between the USA and its European allies, which justifies the visible European reluctance to follow the US lead.

In conclusion, the combined effect of economic crisis East and West is exercising a sobering and positive influence on East-West relations, especially in Europe.

# The European Community and the Newly Industrializing Countries

JACQUES PELKMANS
Maastricht/Florence

**Summary**

The paper investigates the European Community manufactures imports from and exports to 16 newly industrializing countries during the late 1970s at the one-digit sector level. In addition, a disaggregated analysis of the Community's manufactures imports concentrates on possible problem sectors, defined as those having fairly high NIC-import shares in 1975. The interpretation of the nature, speed and selectivity of this trade is based upon a dynamic view of comparative advantages and world-wide relocation of industrial production dependent on the stage of economic development. The forces at work are therefore fundamental and persistent. The appropriate policy response will have to be based upon a well-considered view about how the EC is going to adjust structurally. It is shown that the actual role of the common commercial policy of the EC increasingly threatens to hamper adjustment. The paper concludes with a proposal for a liberal, long-run trade policy *vis-à-vis* the newly industrializing countries, reflecting an awareness of the costs of protection and of shifting commercial power.

The second half of the 1970s has brought a sudden awareness of what was hitherto recognized by only a few. A subset of the "South", called the newly industrializing countries (NICs), are successfully implementing some of the lessons in conventional development economics and comparative advantage, thereby rapidly increasing their competitive exports to the market economies of the North.

The present paper investigates the trade flows during the late 1970s between the European Community and 16 NICs, selected primarily on the criterion of rapid growth of exports of manufactures to the Common Market. This will be done at a fairly aggregate and a disaggregated level of sector specification and considering briefly some other determinants of export increases such as direct investments of Community firms in some NICs and protectionist Community policies. After a short digression on

the relation between trade policy and adjustment to imports of manufactures from NICs, the case for a liberal and well-considered common commercial policy towards the NICs is presented.

## 1. The NICs: who is who?

Although there is no established definition of NICs, one usually adopts the OECD (1979, p. 21) definition: less developed countries (LDCs) that show:

(i) a fast growth in the level and share of industrial employment;

(ii) an enlargement of export market shares in manufactures;

(iii) a rapid relative reduction in the real per capita income gap separating them from most OECD countries.

This delimitation conveys the relevant characteristics, but does not provide strict demarcations for empirical research. As Table 1 shows, NICs are defined quite differently in different studies (see first four columns). Given our concentration on the trade flows between the EC and NICs, and the related trade policy of the Community, it is especially the second OECD criterion that seems to be of primary concern. Yet, in passing, it is not unimportant to note that NICs can be found with widely divergent GNP per capita. Table 1 also shows that the structural growth rate of *total* exports is an unreliable indicator, ranging from 8.8% to 35.2% for 1970-78. In comparison, it should not be ignored that most OPEC countries achieved rates between 24-38% and that many little-noticed newcomers (often with a very small export base) registered export growth rates beyond 20%.[1]

Concentrating on the *composition* of NIC exports, Table 1 brings out clearly that the share of manufactured goods in total exports is either very high for a non-OECD country (South Korea, Hong Kong, Israel) or rapidly increasing. Only two countries out of the 22 in the list (Venezuela, an OPEC member, and Egypt) fail this test and one country is a bit sluggish in this dynamic club (Chile, but political reasons have counted heavily here).

For the purpose of the present paper, concentrating on the Community's relations with NICs, a further delimitation seems appropriate. As elaborated below, an NIC will be defined as a less developed country, the Community's *manufactures* imports from which grow at least as fast as the Community's manufactures imports from the world.

This criterion eliminates those countries that would only be NICs for the USA, Canada, or Japan but not for the EC (like Mexico), as well as countries supplying the EC at a rapidly increasing rate with products other than manufactures, and those with a poor record in export growth of manufactures in general (like Yugoslavia). Since "manufactures" is still an

*Table 1*
*Some indicators of possible NICs (US dollars; %)*

| | Unctad | Edwards | OECD | EC | 1978 GNP per capita | Growth rate of exports: 1970-78 | Share of manufactureds goods in 1968 exports | Share of manufactureds goods in 1977 exports |
|---|---|---|---|---|---|---|---|---|
| South Korea | x | x | x | x | 1 160 | 35.2 | 73.7 | 85 |
| Singapore | x | x | x | x | 3 290 | 24.5 | 22.2 | 44 |
| Taiwan[c] | | x | x | x | 1 400 | | | 49 |
| Hong Kong | x | x | x | x | 3 040 | 19.4 | 94.7 | 94 |
| Mexico | x | x | x | x | 1 290 | 16.0 | 18.2 | 29 |
| Brazil | x | x | x | x | 1 570 | 21.0 | 6.4 | 26 |
| Spain | | x | x | x | 3 470 | 21.0 | 44.7 | 71 |
| Greece | | | x | x | 3 250 | 21.7 | 17.8 | 50 |
| Yugoslavia | | x | x | x | 2 380 | 16.4 | 55.9 | 69 |
| Turkey | | | | x | 1 200 | 13.5 | 3.0 | 25 |
| Portugal | | x | x | x | 1 990 | 11.1 | 59.7 | 70 |
| Argentina | x | | b | x | 1 910 | 16.7 | 10.3 | 24 |
| Philippines | | | b | x | 510 | 15.0 | 7.0 | 25 |
| Malaysia | a | | b | x | 1 090 | 20.0 | 6.1 | 17 |
| Venezuela | | | | x | 2 910 | 18.5 | 0.5 | 2 |
| Chile | | | b | x | 1 410 | 12.1 | 2.8 | 7 |
| Israel | | | | | 3 500 | 19.6 | 70.1 | 80 |
| India | a | | b | | 180 | 16.8 | 44.5 | 56 |
| Columbia | | | b | | 850 | 17.5 | 9.5 | 19 |
| Thailand | | | b | | 490 | 22.6 | 3 | 19 |
| Egypt | | | b | | 390 | 11.9 | 26 | 25 |
| Pakistan | | | b | | 230 | 8.8 | 49 | 57 |
| Ireland | | | | | 3 470 | 20.4 | 32.3 | 55 |
| Japan | | | | | 7 280 | 20.3 | 79.4 | 97 |

*Sources*: Unctad (1979) and World Bank (1980); various tables.
*Notes*: The headings of columns 1 to 4 are: Unctad (1979, p. 24, footnote 3); Edwards (1979); OECD (1979); EC (1979, p. 41, Table III-11, footnote 1). Though Edwards (1979) occasionally mentions other countries, these are not related to general indicators; as his "main" nine NICs are. They are not included above. (a) = At Unctad, 1979, p. 326, this country is added to the six NICs listed. (b) = possible future NICs, or substantial suppliers of manufactures, according to the OECD (1979, p. 24); they are not included in the OECD study. (c) = Taiwan does not always feature any longer in UN official statistics. The World Bank (1980) includes Taiwan. However, note that the manufactures share of exports in 1976 is 85%, according to the World Bank (1979). It seems unlikely that the huge difference is soley attributable to a change in Western protection.

aggregate of very many products of different sophistication, a moderate export performance to the EC in this aggregate may still conceal sharp increases at the product level. The delimitation procedure can therefore only be complete on the highly disaggregated level. Not only would such a complete procedure be extremely laborious, it would also be hard to distinguish incidental EC import jumps from trends. In addition, NICs

tend to spread their industrialization efforts quickly beyond the few initial products and even beyond the one or two starting sectors; to detect patterns therefore one would still need a more aggregate approach.

The above requirement eliminates from the set in Table 1 the following countries, with EC bilateral import growth rates of munfactures below that of the EC import growth rate of manufactures from the world (107.2% for 1975-79): Yugoslavia (86.6%), Mexico (68.1%) Colombia (5.5%), Pakistan (95%) and Chile (86.5%).

It leaves the following 16[2] NICs for the present study: South Korea, Singapore, Philippines, Malaysia, Thailand, Hong Kong, Taiwan, India, Turkey, Greece, Portugal, Spain, Brazil, Argentina, Israel and Egypt.

About half of these countries have special trade relationships with the EC. Greece has become a member of the Community as of 1 January 1981, although its market access will only gradually improve. In a few years, however, its status may be expected to be comparable to another dynamic exporter within the Common Market: Ireland. It is quite obvious that by the middle of the 1980s trade policy measures against Greece will become practically impossible. Spain and Portugal have applied for membership. At present Spain has marginally and Portugal considerably better access than non-Lomé LDCs (as an EFTA member it has free access for most industrial products since 1977, and faced lower tariffs before). Relations with Turkey are theoretically similar to those with Israel but various political sensitivities have caused serious problems in the past (and again in September 1980).

Israel enjoys free-trade-area status for industrial goods since 1977. The Philippines, Singapore, Thailand and Malaysia are members of ASEAN, which has concluded a trade treaty with the EC in 1979.

Its influence on the analysis of the recent past is negligible, but future trade policy will have to reckon with this fact.

None of the 16 NICs considered here are parties to the Lomé Convention, but South Korea, Hong Kong, India, Brazil, Argentina and Egypt benefited from the Community's Generalized System of Preferences, as valid until 1980. Of course, most of the other countries mentioned have this possibility as well. Taiwan, finally, has no official diplomatic status with the European Community any more. Although there are informal contacts it is hardly a speculation to suggest that this leaves Taiwan in an extremely unfavourable position, only mitigated by a relatively large stock of EC direct investments on its territory.

## 2. The NICs as a trade policy issue

If we ignore problems of autocratic domestic politics and of suppressed trade unionism – deplorable phenomena in several NICs (but of course

common to a series of other less-developed countries) – the NICs are a success story. They represent a dream of conventional development economics come true. Therefore, it is not immediately clear what the Community's policy problem could be the NICs would pose. Fast rising GNP per capita and a rapidly increasing working population in manufacturing are typically achievements to be satisfied with. Only the idea that NICs increase their export market shares in manufactures could be worrisome, but it need not be as rapid industrialization creates export markets for the European Community's machinery and technology exporters and induces a general rise of import demand over a large range of products. The mere fact, first brought out by the OECD (1979, p. 8) that import penetration[3] of 10 NICs into the OECD market of manufactured goods in 1977 had barely passed 1% seems to destroy any case for considering the NICs as a threat.

The question is indeed a more subtle one.

The policy problem for the European Community consists in the nature and speed of adjustment to rising NIC shares in a limited number of product markets, both within the Common Market and outside. These problems of adjustment come on top of the strains caused by a less favourable growth climate and by uncertainties about energy prices, inflation and the return to business investment. A major issue of this paper is the role of the Community's common commercial policy in the adjustment to rapidly rising imports from NICs. It will lead us to the question whether an economically and diplomatically superior policy would not be possible.

Whether LDCs are switching from import-substituting to export-oriented industrialization strategies, or development plans are concentrating resources on key products, rather than spreading them out over a host of sub-optimally sized plants in many sectors, initial NIC export strategies tend to be highly specific. Producing large volumes of a very limited range of goods, in which they know they can achieve great cost advantages, they select one big or a few medium-sized OECD economies as export markets. Such precisely targeted export drives in standardized goods may derive from a lack of broad marketing knowledge, but they sometimes seem to be carefully planned within multinational production corporations investing in NICs for the sole purpose of assembly or by world-wide trading companies, distribution chains and international department stores.

In addition, more advanced NICs moving into products with a larger physical capital content may be aided by the local government or, again, by multinational producers to set up large-scale industrial complexes in steel, automobiles or textiles. Once production is on stream, exports may rise extremely rapidly.

The adjustment problem being product-specific and, at that level, one related to sudden jumps in import growth rates, the question is almost by definition of negligible importance on any aggregate level of economic

analysis or political economy. This creates a grave danger for undetected piecemeal protectionism, thought out between specialized bureaucrats and threatened producer lobbies. In the European Community the danger is even greater as external commercial policy is almost completely in the hands of Eurocrats, talking directly with pressure groups, and invisibly supported, or blackmailed, by national bureaucrats in the Committee 113, without a serious parliamentary control, and without publicly accessible procedures or hearings. "Solving" adjustment by piecemeal trade-policy-making on a high level of disaggregation does not arouse a great deal of public interest for perhaps quite a while. A continuous monitoring of it is tiring and boring and hardly makes small headlines even in the best informed newspapers.

*Detecting* the NICs, in so far as they pose a trade policy problem is therefore not quite the same as *defining* an NIC with respect to the aforementioned criteria. On the six-digit product specification level, a less developed country may be seen to have rapidly increasing exports to the EC, or to one or two member countries only, whereas that country's export performance on, say, the two-digit sector need not show up as above-average. It might also fail to classify as an NIC on the basis of rising GNP per capita or on the basis of an increasing share of employment in manufacturing. It follows that, whilst the definition of an NIC takes place at the aggregate level very similar problems may also be caused – on the disaggregated level – by generally unsuccessful, poor countries with a single products line that manages to penetrate the EC. The point is that the NIC problem for the EC (and the OECD in general) is inevitably part of the overall adjustment to shifting comparative advantages in standardized products to all less developed economies.

More precisely, it has to be placed in the framework of dynamic comparative advantage to acquire an insight in the fundamental and persistent forces at work. Though the economics of these processes are exceedingly complex, their broad nature is well-understood and has been given empirical content (Balassa, 1979).

The idea is that economic development in a fairly open world economy tends to follow a stages approach with respect to factor input and types of products competitively produced for the world market. On the factor input side, availability of a minimal physical, social and educational infrastructure is required. This has proved to be easier in development plans than in actual practice, especially with respect to the latter two. Without these infrastructural provisions it is difficult to expect that product and factor markets can work in a poor country. Moreover, it seems ever more difficult to achieve stability of supply and expect sufficient productivity – even with the lowest wages – to result in cost advantages over established world producers. Typically, today's NICs tend to have either a high literacy rate or an ample supply of literate, but low-skilled

workers (the latter applies to India and Egypt, for instance).

The types of products at the lower stages of development and comparative advantage can be associated with small-scale manufactures production like clothing of simple varieties, simple non-fashion footwear and handbags, simple toys and souvenirs and certain simple processed primary commodities, statistically counted as manufactures, such as certain wood manufactured products. The emphasis on "simple" and "small-scale" has to do with the presumed shortage of physical capital, needed for large-scale production and the presumed shortage of entrepreneurial, organizational, design and marketing capacity in countries that begin to industrialize.

They may advance by gradually building up experience, purchasing technology, or machine-embodied technology so as to diversify into more demanding products. The entry into higher stages may be accelerated by foreign direct investment bringing in excessively scarce factors. These investors may reorganize or create production but they may also facilitate direct entry into markets of advanced countries by subcontracting or production for retail chains.

The other, not mutually exclusive, route is stringent planning of industrialization by concentrating resources on a few key products that are sufficiently standardized to expect export success, yet sufficiently advanced to move beyond the beginnings of efficient manufacturing. The "rifle-shot" approach[4] makes sense in avoiding sub-optimal plant size but obviously risks a collision course in willingly pursuing the disruption of advanced country markets if it is done too aggressively. However, "rifle-shot" incidents seem to be typical of young NICs with insufficient diversification.

Further stages would push the now semi-developed country into less standardized products, or physical capital intensive goods with simple production processes while perhaps still improving on its productivity in the low-skill intensive, standardized goods in order to continue competing with emerging young NICs. Eventually the NIC "graduates" into something like a newly advanced country (an NAC?) with a broad industrial base, possibly seeing its initial success sectors decline into comparative disadvantaged sectors and still aiming for the production of certain specialized high technologies.

Of course, one should not exaggerate the model-like smoothness of this complex process. One should also beware of the fiction that industrializing LDCs are all strongly export-oriented. Many of them are not. Some are basing their development strategy on exports in a few key sectors only. It is a minority that is truly export-oriented. The trade policy issue is of course not that poor countries are industrializing but that a small but growing number of them pursue the objectives of productive efficiency, growth, urban employment and maximizing foreign exchange earnings by relying on exports of manufactures.

The *economic* case for singling out NICs as a subset is then that the combination of product specificity and extreme import growth rates can cause severe problems for a limited group of firms and traders that might push up the cost of immediate adjustment beyond the society's gains from trade.

The *institutional* case for singling them out is that GATT, Art. 19. dealing with "market disruption", is often avoided and obscure, bilateral, sometimes even unpublished trade policy instruments are applied as substitutes. It is the NICs that suffer most from these "lawless" commercial policy dealings, only too often ruled by power and too rarely by considerations of industrial adjustment. As demonstrated by the Community's feverish attempts to obtain agreement on a "selective safeguard clause" in the Tokyo Round, the NIC issue has now assumed great institutional significance.

## 3. Recent EC-NIC trade in manufactures

Within the European Community and beyond, protection against present NICs and old ones like Japan has often been defended in trade policy circles by pointing to extreme import growth rates as well as to "intolerable" bilateral trade deficits. The first phenomenon is suggested to indicate market disruption and the second one an inappropriate "burden" of adjustment. As both indicators will be used in this section, some further discussion seems desirable before examination of the evidence.

In the first place, import growth rates may be considered as a signal that possibilities of "market disruption" might arise especially in times of recession. However, a proper investigation would have to be conducted at a high level of disaggregation. Moreover, the case should be made that "serious injury" is attributable to the imports and not to technological change, factor substitution (i.e. replacing men by machines) or a fall in demand. Usually, though, it is one or more of the latter factors which cause a fall in the "injured" subsector's employment. If so, and production capacity does not fall with employment, firms may be confronted with decreasing market shares due to (NIC) competition. Such developments create strong pressures for protection because labour union and management can jointly lobby, the former primarily to protect jobs in that sector and the latter primarily to protect market shares and profits. Of course, lobbyists must present their particular case as one in which adjustment is impossible in such a short time-span and undesirable given the sectoral unemployment. Import growth rates conspicuously higher than domestic production help to make the case for protection politically digestible, especially if it is directed at the "culprits" only.

In the second place, bilateral, sectoral trade balances over time may be

considered as a signal of declining competitiveness in the product group concerned. But there is nothing wrong with declining sectoral competitiveness *vis-à-vis* certain countries, here NICs, as long as competitiveness improves in other product groups. Economic growth in open economies such as the EC is precisely a result of continuous shifts to sectors of higher productivity either by further specialization within a sector or by intersectoral reallocation of productive resources. Furthermore, concern with bilateral trade flows in a multilateral trading (and payments) system is alien to its nature. Disaggregating them further into product categories, even if only of the one-digit variety, makes worries about balanced flows even less warranted. This holds especially in a North-South context where there seems to be ample scope for inter-industry specialization (rather than intra-) necessarily creating disaggregated "imbalances" in trade flows. Such "imbalances" are precisely the result of the deepening world division of labour. Finally, one might fear overall trade imbalances for the EC because of a loss of competitiveness.The implicit assumption is then that exchange rates do not change. If exchange rates could adjust, however, there will always be *comparative* advantages that can be exploited. The conclusion is that sectoral, bilateral trade imbalances over time can be utilized as indicators for competitiveness. If such imbalances enter as arguments into commercial diplomacy and if access to both markets is roughly comparable, there is no economic justification. One suspects that the political motivation could derive from the desire to "do something" about NIC competition or from a neo-mercantilist preference to maintain certain industries, without immediately slapping tariffs or "voluntary" export restraints on their supplies. The imbalance argument has been a typical prelude to further protection in the case of an experienced ex-NIC, Japan.[5]

Pleas for protection tend to rely on other economic objectives (sectoral employment, preferably at the going wage rate) or political objectives (votes; minimum sector size for "security") than real income growth. A further essential point is that protection imposes costs on the domestic consumer *and* the foreign (NIC) firms, especially on workers, that only rarely enter the calculation of decision-makers. Would policy-makers in advanced countries *explicitly* underwrite their *implicit* argument that a worker in a poor NIC, thrown into unemployment because of protection in export markets, can "adjust" with less injury than a worker in a Welfare State, thrown into unemployment because of NIC competition?

Table 2 provides data about the recent developments in the Community's trade in manufactures with the 16 NICs selected. The information is presented in such a way that some major trade policy indicators (economically justifiable or not) can be deduced from it. Essential to the NIC phenomenon are import growth rates, of course. However, such growth rates have to be used with a least some information on the import value in the base-year (1975). Whenever the import base in 1975 is low or very

*Table 2*

*EC-NIC trade in manufactures: 1975-79*

(growth rates in %; millions of ECU)[1]

| Imports from / Export to | SITC[4] 5-8 (%) | | SITC[4] 5 (%) | | SITC[4] 7 (%) | | SITC[4] 6/8 (%) | | Net EC exports[2] of manufactures (1979; in ECU) |
|---|---|---|---|---|---|---|---|---|---|
| | Imports | Exports | Imports | Exports | Imports | Exports | Imports | Exports | |
| World | 107 | 84 | 120 | 103 | 108 | 74 | 105 | 88 | 65 563 |
| All NICs | 166 | 70 | 118 | 87 | 236 | 62 | 153 | 74 | 8 885 |
| South Korea | 210 | 171 | 51[a] | 190[a] | 459[a] | 149 | 190 | 292[a] | −415 |
| Singapore | 156 | 108 | 186[b] | 153[a] | 178 | 98 | 132 | 110 | 480 |
| Hong Kong | 125 | 188 | 58[b] | 173[a] | 196 | 181 | 117 | 198 | −1 161 |
| Taiwan | 175 | 101 | 215[b] | 177[a] | 315[a] | 75 | 144 | 103[a] | −748 |
| Thailand | 377[a] | 119 | 378[a] | 184[a] | 348[b] | 102 | 378[a] | 99[a] | 298 |
| Malaysia | 182 | 94 | 66[b] | 140[a] | 279[a] | 88 | 161 | 85[a] | 282 |
| Philippines | 697[a] | 97 | negl.[3] | 119[a] | 1 028[b] | 85 | 663[a] | 128[a] | 286[c] |
| India | 146 | 116 | 119[a] | 56 | 110[a] | 68 | 149 | 243 | 549 |
| Israel | 164 | 39 | 106[a] | 106 | 178[a] | 33 | 180 | 31 | 846[c] |
| Egypt | 710 | 95 | 110[b] | 27 | 2 410[b] | 148 | 650[a] | 62 | 1 740[c] |
| Turkey | 133 | −16 | 58 | 33 | 143[b] | −22 | 138 | −33 | 978[c] |
| Greece | 141 | 87 | −12[a] | 92 | 41[a] | 60 | 161 | 139 | 1 913[c] |
| Spain | 188 | 60 | 195 | 86 | 257 | 51 | 148 | 58 | 452 |
| Portugal | 110 | 96 | 103[a] | 129 | 87 | 103 | 117 | 69 | 751 |
| Brasil | 169 | 6 | 77[a] | 52 | 329[a] | 3 | 143 | −16 | 1 355[c] |
| Argentina | 234[a] | 117 | 93[a] | 82 | 174[a] | 160 | 354[a] | 73 | 1 279[c] |

*Source*: Calculated from: Eurostat, *Monthly External Trade Bulletin*, various supplements.

*Notes*: [1] For 1975, read EUR for ECU. In 1975 1 EUR = (approx.) $ 1.24; in 1979, 1 ECU = $ 1.37 (average used by Eurostat).

[2] Net exports = exports minus imports. A negative sign indicates a (bilateral) deficit for the EC (in manufactured goods only).

[3] negl. = negligible absolute import volume throughout the period.

[4] SITC 5 = chemicals; SITC 7 = machinery and transport equipment; SITC 6; 8 = other manufactured goods.

[a] 1975 im/export base below 100 million EUR (low base).

[b] 1975 im/export base below 10 million EUR (exceedingly low base).

[c] EC bilaterial surplus is larger than the EC imports from that country (in manufactures).

low, thereby enabling magnified growth rates that need not be alarming at all, this has been indicated. Policy-makers also frequently use the bilateral trade balance as an indicator. Though its use as an indicator is often suspect, it has nevertheless been provided as well. Finally, bilateral EC export growth rates are given.

The picture that arises is clear. The Community is increasing its imports of manufactures from the NICs at a speed two-thirds higher than the rate of growth of manufactures imports from the world as a whole. At the same time its export performance in the same goods to NICs is markedly inferior to the overall growth of manufactures exported to the world. However, the 1979 surplus with the NICs in these goods is quite comfortable by any standard: close to 9 billion ECU, which is 48% of the Community's entire imports of manufactures from the NICs! But the sharply diverging growth rates of exports and imports, if continuing, are bound to undermine this comfort rapidly. The second half of the 1970s surely points to a worrying trend: not only did the absolute EC surplus in manufactures trade with the NICs decline from 9 174 million EUR to 8 885 million ECU,[6] but the 1975 surplus was 32% *larger* than all EC imports of manufactures from NICs whilst the 1979 surplus amounted to a mere 48% of those EC imports.

Disaggregation into three one-digit categories shows that the true NIC problem should not be looked for in chemicals (SITC 5; SITC refers to Standard International Trade Classification). Not only are EC chemicals imports from all NICs growing a bit more slowly than EC chemicals imports from the world, the country pattern is extremely uneven and the 1975 base is very often low or extremely low. Moreover, the share of chemicals in total EC NIC-manufactures imports declined from an already small 5.3% in 1975 to 4.4% in 1979. Chemicals will therefore be ignored below[7] as a comparative disadvantage good for NICs.

For machinery and transport equipment the rate of import growth is very high (236%), enlarging its share in EC NIC-manufactures imports from 18% in 1975 to 23% in 1979. For a number of NICs, their base value of exports to the EC is low or extremely low, so that the sometimes fantastic EC bilateral import growth rates will still mean very little in terms of adjustment. This is reflected in the EC surplus in these goods: the huge difference in import and export growth rates has brought the surplus down from six times the absolute imports from NICs in 1975 to a little over twice the absolute imports from NICs in 1979, still very large though. The absolute surplus even increased from 7 324 million EUR to 9 659 million ECU. Furthermore, in 1979 no less than 46% of EC NIC-imports of machinery and transport equipment originated from Spain alone. The fact that these large imports have been growing some 257% over 1975-79 seems to have been much more important than sensational growth rates of initially negligible imports from many other NICs.

It is often thought that machinery and transport equipment represents a category of goods strongly competitive *vis-à-vis* LDCs, even NICs. It appears, however, that the European Community's surplus in machinery and transport equipment *vis-à-vis* the NICs during the second half of the 1970s conceals a decline of comparative advantage. Even if "comparative advantage" is a meaningful concept at such a high level of aggregation – I would submit that it is not – one has to qualify such a sweeping conclusion by looking more closely at Table 2. Of the 16 NICs only three show both a non-small export base in these goods *and* a high export growth: Spain, Hong Kong and Singapore. These three countries have a per capita GNP of more than $ 3 000 (1978)[8] and may not be fully comparable with most other NICs with respect to the requirements for exports in these semi-advanced products, but rather with Ireland, or perhaps even Italy. They seem to confirm the notion that NICs pose different competitive challenges at different layers of development, both because of product-wise and sectoral diversification on the output side and because of improved endowment of human, physical and technological capital on the input side. Unfortunately, the data in Table 2 are too aggregated to enable a proper judgment on this matter, since SITC 7 lumps together goods with very different input requirements and imports of widely diverging degrees of standardization.

Most of the concerns on the NIC exports to the EC stem, however, from imports of "other manufactured goods" (such as textiles, steel semi-manufactures, furniture, clothing, footwear, etc.). Even at this aggregate level there are a series of indications for the competitiveness of the NICs. All 16 NICs individually show higher export growth rates (to the EC) than the world does; the four NICs that had a 1975 export base of below 100 million EUR have very high growth rates indeed (354-663%); the five biggest EC suppliers of these goods among the NICs increased their exports to the Community by 139% despite the dampening effect of (increasing) protection, often especially applying to their shipments; the small deficit of the Community with the NICs in these goods in 1975 had increased tenfold by 1979 and had risen to 39% of EC imports from the NICs.

But even here disaggregation seems warranted before jumping to conclusions. If the EC is believed to be uncompetitive in these goods, how can one understand the fairly strong Community export performance to several East Asian NICs and to India in these goods?

A different picture of the Community's NIC manufactures imports is presented in Table 3, where the growth in NIC shares in EC imports of manufactures has been provided for the Nine as a whole and for seven of its eight economies.[9]

In machinery and transport equipment all NIC shares in EC imports increase but none of them had even reached the 7% level in 1979. In other

*Table 3*
*NIC shares in manufactures[1] imports of the European Community*
(1975-79; %)

| | SITC 7 | | SITC 6/8 | |
|---|---|---|---|---|
| | 1975 | 1979 | 1975 | 1979 |
| EC[2] of 9 | 2.70 | 4.36 | 8.60 | 10.71 |
| FR of Germany | 4.0 | 4.9 | 11.1 | 13.1 |
| France | 3.9 | 6.4 | 7.3 | 9.7 |
| Italy | 1.7 | 6.3 | 7.0 | 11.2 |
| The Netherlands | 1.9 | 2.9 | 5.0 | 7.6 |
| Belgium/Luxembourg | 0.9 | 1.0 | 4.6 | 6.7 |
| United Kingdom | 2.7 | 4.3 | 12.7 | 13.1 |
| Denmark | 1.0 | 1.6 | 5.9 | 2.3 |

*Source*: Calculated from source, Table 2.
*Notes*: [1] Only SITC 7 ("machinery and transport equipment") and SITC 6/8 ("other manufactured goods"); not chemicals (SITC 5).
[2] EC includes Ireland; Ireland is not listed separately.

manufactured goods a similar growth picture arises[10] although the NIC shares are much higher.

This begs the question how arbitrary it is to use a threshold NIC share of EC imports beyond which NIC competition can be considered to become "important" or "sensitive". First of all, NIC shares may rise due to competition among non-EC suppliers without necessarily increasing market shares. In the period considered this is not the case.

Not only the NIC share of extra-EC imports has risen over the period both for machinery and transport equipment (from 8.1% to 12.6%) and other manufactured goods (from 21.0% to 25.0%), but the ratios of extral over intra-EC imports have moved up as well (respectively from 0.50 to 0.53, and from 0.69 to 0.74). In the second place, a rising import share accompanied with rising home production for exports, say in quality goods, need not inflict any pressure on domestic producers, facilitates the shift to higher value-added output and is beneficial for consumers. Therefore, what one really wants to know are "market shares" (or the rate of import penetration) rather than "import shares". The import penetration rate, defined as imports (from NICs) as a percentage of apparent domestic consumption (domestic production minus exports plus *all* imports), is very hard to come by for statistical reasons.[11] This is unfortunate since a rising penetration rate necessarily means more competition and a lower market share in the domestic market. The loss of strength in the domestic market is a fate every producer would like to avoid, no matter how strong his exports are. Furthermore, competitive pressures of NICs are easily relayed to third markets which would spell contraction abroad. However, the

condition for using penetration rates and even more for import shares, to indicate the "sensitivity" to NIC imports is that the level of disaggregation is high so as to be reasonably sure about the homogeneity of the products. But even at the three-digit product specification of many products this homogeneity condition is not easily fulfilled. A striking illustration of the point is in SITC 831 (travel goods, handbags, etc.) where Italy joins other EC members in having high NIC shares of these imports while persisting its export drive unabatedly, in contrast to all other EC members where exports have vanished. The explanation is that further disaggregation would show Italy's great comparative advantage in the luxurious "up-market" products within this category. Using import shares is therefore second best to penetration rates, if available at all. The method is not fully reliable as an indicator of "sensitivity" to NIC competition.

It follows that Table 3 should be used with caution. Perhaps the most interesting aspect is the *change* in import shares. In machinery and transport equipment Spanish auto exports to France and Italy seem to be a major cause for a rapid increase of the NIC share, while in other manufactured goods the same phenomenon is more muted. Another interesting feature is the moderate increase in the NIC share of British imports of "other manufactured goods". One wonders whether Britain has been allowed to be more protectionist, either informally or within the common commercial policy.

## 4. A disaggregated analysis of selected sectors

The previous section provides a useful perspective but the core of the NIC problem can only be studied by disaggregating further (Hager, 1980). On the other hand, it seems impossible to extract a coherent picture from the multitudinous nitty-gritty issues of daily trade policy on the product level.

A reasonable compromise will be adopted in taking a considerably higher level of disaggregation than in the previous section without risking to lose oversight by studying many thousands of single products. This can be done at the intermediate level of SITC specification. The example of textiles (not clothing) may illustrate what is involved. Textiles are first specified as a sector at the two-digit level (as 65): the NIC share in total EC imports of textiles was 9.3% in 1975. At the disaggregated level one may observe an NIC share of 1.5% in EC imports of "yarn of continuous synthetic fibres, etc." (SITC 65161), next to an NIC share of 58% in "cotton yarn and thread, grey, not mercerized" (SITC 6513) and one of 27.9% when the same cotton yarn and thread is mercerized, bleached or dyed (SITC 6514). Yet, the EC total of the latter two imports together, though on the four-digit level, is only half the value of the EC total of the imports of the former, despite its classification at the five-digit level.[12]

Except for specific sectoral studies, it seems more useful to move beyond the casuistry of literally thousands of product types with very different import weight and attempt to grasp the nature and extent of the problem of varying levels of intermediate aggregation. Even with such an approach the degree of detail already hampers the consistent reference to the underlying process of world-wide structural economic change.[13] It also has to be realized that the choice of "machinery and transport equipment" plus "other manufactured goods" makes the analysis even more selective.

One drawback of this method ought to be kept in mind as it occasionally may play a role: at the intermediate levels of disaggregation, products are classified with respect to *end-uses* but these need not imply similar *input* requirements. In the framework of dynamic comparative advantage information on both is required. In terms of our textile example, "yarn of continuous synthetic fibre" (SITC 65161) is essentially a chemical product from a physical capital-intensive industry and often based on oil derivatives, whereas "cotton yarn and thread, grey, not mercerized" (SITC 6513) can in principle be produced on simple spinnning machines with substantial low-skill labour, raw cotton and small-scale plants. The great difference in NIC shares of EC imports of the two textiles can be readily explained in terms of availability of factor inputs if one disaggregates fully, whereas some of these insights are lost when imports are studied at the level of "textile yarn and thread" (SITC 651).

Table 4 provides summary information by means of various indicators on the NIC performance in EC manufactures imports (except chemicals) in a number of selected three-digit sectors, with some-two or four-digit ones if relevant. Sectors have been included if, in 1975, the NIC share of EC imports of the category concerned was above 10%. If this was the case already at the two-digit level, higher degrees of disaggregation have been avoided. In other words, the highest degree of aggregation has been chosen at which the 1975 NIC share in sectoral EC imports was above 10%. It is expected that, with the possible exception of a single product case outside these sectors, a fairly broad, yet disaggregated picture of NIC-EC manufactures trade, and particularly its sensitive elements, can be so provided. However, one should appreciate the selectivity of the presentation in its concentration on possible problem sectors. In addition, two exceptions to the above-10% criterion have been made in cases where evidently no adjustment issue has arisen in the European Community.[14]

Eleven out of 13 sectors show an enlargement of the NIC share in EC imports of more than one percentage point, with radio receivers and travel goods, handbags, etc., making a spectacular jump of some 14 points in four years only. Calculating and accounting machines is the only sector where the NIC share declined (by 1.7 points) but even here one has to be careful about the volume as the simple calculators declined in price (at least relatively to the more sophisticated ones). NIC exports in this sector are

heavily concentrated in simple calculators the production of which has now become fully standardized.

The general NIC performance in what presumably could be "problem sectors" is remarkable. As we shall see later, in a number of these sectors NICs suffer from protection, often quantitative protection. Table 4 strongly suggests that such protection has served as a brake at best.

Four sectors have moved from NIC shares in EC imports of above 20% to shares above or equal to 30%. This seems to be rather high and may be expected to lead to adjustment problems if import value or speed (or both) are considerable. The 1979 import values (from NICs) are surely not small ($ 372 million or beyond) with clothing as a very large one. The rates of import growth are much higher than one might expect by comparison to non-problem sectors or overall rates. For example, all four are considerably higher than the overall EC import growth rate for SITC 6/8 (105%; see Table 2) or than for chemicals (120%). Another interesting comparison is with the 6% growth norm of the first Multi-Fibre Arrangement (1973-77): if these high NIC share sectors would have been subject to an across-the-board rule of 6% EC import growth from NICs, and if EC inflation in the period is assumed to have been 10% annually,[15] the import growth rate for NICs would have been 81%. The four growth rates, however, are close to or beyond twice such magnitudes.

If one pays attention to import growth rates – a crucial indicator for NICs compared with other LDCs – the performance is impressive, given increasing protection and a low growth climate in the Community. Assuming an EC inflation of 10% annually, the real average import growth for 11 out of 13 sectors has been above 14% *annually* (from NICs) and even reaches to slightly beyond 25% (travel goods, handbags, etc.). Though Table 4 gives a very partial view of the Community's imports, it does show that the NIC performance has now spread over a considerable number of sectors combining fairly high import shares for these countries with high growth rates of their supplies.

On the other hand, the competitive pressures from NICs are still heavily concentrated in these 13 sectors. The 10 sectors resorting under "other manufactured goods" make up no less than 74.6% of the total EC NIC-imports in these goods in 1979, whereas the three sectors listed, resorting under "machinery and transport equipment", only reach 22.5% of total EC NIC-imports of these goods (which are already modest; see Table 3). Apart from radio and TV sets and perhaps some electrical appliances, SITC 7 does not encompass adjustment problems *vis-à-vis* the NICs. Even in the products mentioned, it is Japan which causes pressures, when they arise at all. This holds especially for TV sets and not so much for radios although the latter's EC import growth rate is much more impressive and starts from a larger base. The big electric/electronic companies of the Community seem convinced that radios are a market without

*Table 4*

*Some characteristics of EC[1] imports from NICs; selected sectors*

*(1975-79; %; millions of US$)*

| Sector | SITC[2] | | EC 1979 import from NICs | Share in 1979 EC man.s imports from NICS (%) | | Import growth rate 1975-79 | NIC share in EC imports of sector | |
|---|---|---|---|---|---|---|---|---|
| | old | new | $ | SITC 6/8 | SITC 7 | (%) | 1975 | 1979 |
| 1. Veneers, plywood | 631 | 634 | 538 | 4.0 | | 261 | 14.6 | 21.8 |
| 2. Textile yarn and thread | 651 | 651 | 1 015 | 7.6 | | 171 | 12.9 | 17.0 |
| 3. Woven cotton fabrics | 652 | 652 | 552 | 4.1 | | 145 | 15.2 | 15.8 |
| 4. Made-up textile articles | 656 | 658 | 420 | 3.1 | | 151 | 26.8 | 30.0 |
| 5. Cutlery | 696 | 696 | 90 | 0.7 | | 95 | 18.0 | 19.8 |
| 6. Metal household articles | 697 | 697 | 248 | 1.8 | | 185 | 17.1 | 18.3 |
| 7. Calculating and accounting machines | 7 142 | 7 512 | 113 | | 2.7 | 0 | 19.3 | 17.6 |
| 8. TV receivers | 7 241 | 761 | 248 | | 5.8 | 177 | 7.5 | 14.5 |
| 9. Radio receivers | 7 242 | 762 | 594 | | 14.0 | 179 | 21.9 | 35.3 |
| 10. Travel goods, handbags | 83 | 83 | 372 | 2.8 | | 335 | 22.9 | 36.8 |
| 11. Clothing | 84 | 84 | 5 291 | 39.4 | | 141 | 27.3 | 30.5 |
| 12. Footwear | 85 | 85 | 894 | 6.7 | | 205 | 16.0 | 19.6 |
| 13. Toys, games, sport goods | 894 | 894 | 589 | 4.4 | | 192 | 18.2 | 21.3 |
| Total | | | | 74.6 | 22.5 | | | |

*Source*: OECD, *Trade by commodities*, Market summaries, Imports (2 Vols), Series C, 1975;
OECD, *Microtables import/export 1979, by country*. Calculations by the author.

*Notes*: [1] EC does not include Ireland.
[2] The sources of the data base (see below) have switched to the SITC, second revision, during the studied period.
[3] Only SITC 6/8 and SITC 7; not chemicals (SITC 5).

any growth potential and an evident comparative disadvantage (only France has had temporary protection). The colour TV set market is not yet saturated, however, commits much more resources than radio production (per set) and pertains to some significant technological issues. The most important one is patent protection for PAL colour technology which, it is feared, cannot compete with Japanese technology once the patents expire (1981). The significance of the NICs is merely that they further add to these pressures. In mid-1980 one could observe Philips (Dutch) and Thomson (French) trying to organize a strong case for EC trade protection against colour TV sets to replace the shelter provided by the patents.

Another example in SITC 7 can clarify why the NICs only exceptionally cause sectoral adjustment problems there. The dominant share of Spain in the NIC share of imports of "machinery and transport equipment" is not due to the sectors mentioned in Table 4 but to such sectors as "office machines other than calculating and accounting machines", certain "non-electrical machinery and parts", "internal combustion engines" and, above all, automobiles, primarily exported to France and Italy. In all these internationalized sectors, the NIC (or Spanish) share is very small for the Community – except for French auto imports from Spain. Moreover, trade often results from direct investments of Community firms so that the adjustment is taken care of by the market. However, the small NIC share in *total* SITC-7 EC imports of these goods still represents a large part of the EC imports of "machinery and transport equipment" *from NICs*.

The "machinery" sector seems to be one of the few sectors where EC direct investments in NICs generate substantial imports of goods previously produced in the Common Market.[16] Trade-creating direct investments tend to accelerate the exploitation of the potential comparative advantages of NICs. A dramatic instance can be detected in the case of thermionic valves, etc. (SITC: old, 7293; new, 776) where the 1975 NIC share in EC imports is only 6.2%. But this does not include Dutch non-allocated imports – presumably from Philips's subsidiaries[17] in Taiwan and Singapore –, bringing the NIC share to 19.0%! In 1979, the NIC share has grown to 23%, including 11.0% or close to half a billion US dollars for Philips alone. It is typical that the extremely internationalized electronics sector plays the game of protection (Philips on colour TV sets) together with the exploitation of comparative advantage of NICs in specific other goods (Philips in the Far East) and research-intensive breakthroughs in advanced technology (Philips on video/disc innovations).

The story is quite different in "other manufactured goods". The nature and extent of the adjustment problem is not merely determined by the combination of initially high import shares of NICs (often higher than in SITC 7) with high growth rates of NIC imports, but also by the fact that further specialization into high value-added "up-market" products has

become increasingly difficult because entire sectors have lost their comparative advantage. As is known, the adjustment pains of dismantling and dismissals so as to move to another sector (intersectoral adjustment) are usually considered to be much greater. For workers of any skill-level except the minimum, it may imply that their sector-specific knowledge and experience depreciates sharply in value, in turn causing personal problems of acceptance and lower wages elsewhere. Even more serious is the possibility that the skill profile of workers moving out of one sector is so different from that wanted for the vacancies in expanding sectors that frictional unemployment becomes structural. Finally, the overall climate of stagnant consumer demand and falling investment in the Community, on top of business insecurity about future energy prices and the borrowing capacity of important purchasers of EC exports (such as some LDCs, a few NICs and Poland, for instance) is not particularly conducive to smooth shrinking processes in comparative disadvantage sectors.

This has led to increasing pressures for protection especially in "other manufactured goods", mostly against NICs but sometimes also against the USA (in steel, synthetic fibre carpets) and others. Table 4 has to be considered with at least a global notion of the protection involved. This paper is not the place to deal with actual protection in depth. The extent and coverage, let alone the incidence, of today's protection is exceedingly hard to capture. The following evidence claims little more than illustrative information on some of the sectors listed in Table 4. Firstly, some indications of actual tariff protection can be provided. In an interesting paper Olechowski and Sampson (1980) have calculated the weighted average tariffs if one does not only look to official tariff schedules listing MFN tariffs,[18] but also takes into consideration that the EC has a plethora of special trade agreements, including, of course, the Generalized System of Preferences, all lowering tariffs for the trading partners concerned. Though sectors are not identically defined as in Table 4, the impression that EC tariff rates for developing countries are very low is confirmed: wood and wood products, 1.6%; textiles and textile arts, 8.9% (but only 4.2% for developed trading partners); footwear and accessories, 10.1% (and only 3.3% for developed countries) and miscellaneous manufactured articles, 9.5%. Surely, these low tariffs even in sensitive goods cannot be a true hindrance for NICs.[19]

Secondly, a notion of various non-tariff distortions can be obtained by studying *official* documentation (note that this approach merely establishes a minimum estimate of non-tariff protection). Here, the picture changes dramatically. Olechowski and Sampson apply a frequency index indicating the (unweighted) share of the number of classification headings of goods on the four-digit level the imports of which are subject to some form of import control. The controls may be little more than early warning systems such as surveillance schemes (the EC had one for cutlery, a sector listed in

Table 4); others could be quotas, licensing arrangements, import price controls (as in steel; at present only relevant for Spain and South Korea) and various officially agreed export restraints. Non-tariff distortions of the EC turn out to be very frequent and 90% of them are discriminatory. The share of discriminatorily "controlled" trade items in the total number of four-digit items for "wood and wood products" is 18.4%, for "paper" (not listed in Table 4) 27.7%, for "textiles and textile arts" 75.4%, for "footwear" 15.7% (which is one-third of the US rate) and for "basic metal products" (like cutlery) 7.9%. Miscellaneous manufactured articles have, however, only 3.7% of items "controlled" (all figures for 1976). Other work confirms the sharp rise in EC protection *vis-à-vis* developing countries.[20] According to Page (1980), 34% of EC manufactures imports from developing countries in 1979 is, what she calls "mainly managed",[21] which is higher than the OECD overall share.

Thirdly, there are the two Multifibre Arrangements, preceded by the Long-term Arrangement on Cotton Textiles. Over time these arrangements have become broader in product coverage and in coverage of "fibre" exporters (now 29 countries!), more stringent in terms, permit many more downward exceptions to the rule of 6% import growth and leave ever more diplomatic room for bilateral dictates.[22] In textiles and clothing the EC has now become a champion in protection. Therefore it is truly remarkable that the four sectors listed in Table 4 still register import growth rates of 141-171% which is approximately double the 6% rule (including 10% average annual inflation). The explanation is to be found in the chances given to younger NICs facing less stringent controls than South Korea, Hong Kong and especially Taiwan in clothing and the relatively mild attitude towards Spain and Portugal, and a few young NICs in certain textiles. In textile yarn and thread, NICs like Turkey, Egypt, Argentina and Israel managed to increase their exports to the EC by 400% or more. In woven cotton fabrics it is especially Greece, Spain, Egypt, India and Thailand; in made-up textile articles it is South Korea, Portugal and Israel; in clothing, it is Greece, India, Thailand, Singapore and the Philippines. Established suppliers, on the other hand, see their exports increase by barely more than inflation.

In the case of textiles it is hard to tell whether we confront an NIC problem or a more general North-South problem. The essential distinction between the two is the import growth rate, once a substantial import level has been achieved. One gets the impression that the second, and the soundings about the third Multifibre Arrangement are basically attempting to maintain market shares for domestic producers. Hence, the concern is *not* about the rate of import growth, disrupting markets and seriously injuring local (EC) producers, but about the *level* of imports. The forms of protection chosen – quantitative, mostly – seem to hint at such an inference as well. But that would point to very different underlying causes for protection.

Concern about the *level* of competitive supplies can be translated into concern about the levels of profits and wages, or about employment if capital/labour substitution would be shown to be irrelevant and NIC imports crucial. Hence, if we accept the results of many studies on the impact of LDC exports on employment[23] in the OECD economy (shown to be minimal), concern about import levels, and the resistance to even moderate increases of export shares of poor countries – NICs or not – point to issues of (domestic, not global) income distribution. In contrast, the NIC problem ought to be strictly related to the social, human and technical processes of adjustment that take time and should not become more costly than the gains from relying on cheap imports.

On the other hand, Table 4 and the material underlying it show clearly that even in textiles and clothing younger NICs still have switching possibilities that are not feasible for undeveloped countries with incidental clothing firms. It is the "graduating" NICs the textiles and clothing exports of which are throttled.

## 5: Adjustment to NICs and the Community's trade policy

The European Community, being a set of economies of the mixed economic order, may choose between three distinct methods of adjustment to manufactures imports from NICs: by market forces alone, by industrial policy or by trade policy. Although current practice clearly shows a reliance on some mixture of the three, it is useful for purposes of analysis to establish to what extent they are substitutes. In what follows, to be a substitute will not only refer to the *effects* upon (the direction of) industrial structural change, but also to the level and allocation of the *costs* of adjustment. In turn, these properties can be traced back to political processes in mixed economies, both nationally and at the Community level.

Given certain conditions, market processes lead to effective, early and continuous adjustment. In principle, markets are signalling devices and ignoring signals of declining sales or profits is penalized after some time. The effects on industrial structural change will be to have resources shift out of comparative disadvantage products into similar but higher value-added goods (existing, newly invented or differentiated) or into less similar, if not entirely different products. Specialization within the same industry is normally considered to be a less costly form of adjustment than shifting resources between industries. It is often claimed that market integration within the EC in the 1960s has been easy since it took the form of intra-industry specialization in a large number of sectors. In an open world economy with unabated technological progress, the pressures

for continuous adjustment to change are substantial and their origins complex. Markets simplify the monitoring of change by coordinating information and transforming it into signals that consumers, retailers and producers immediately understand.

A non-exhaustive list of conditions for markets to perform their tasks properly includes workable competition, reasonable predictability of demand and supply trends (or at least market agents ought to have that perception) and a fair degree of inter-regional and intersectoral mobility irrespective of age, of skill, and of cyclical downswings. If these conditions are approximated, market processes tend to minimize the costs of adjustment for the factors of production involved, tend to allocate these costs to the "adjusters" and not to others, and tend to create, by rewarding efficiency and superior performance, an incentive system for the persistence of competition and for new entrants (from anywhere).

This is essentially how market adjustment processes have functioned when national EC markets were opened up around the turn of the 1960s. When the Dillon and Kennedy Round, as well as a quantum jump in international direct investment, further internationalized the economies of Member States, market processes were still treasured as an effective, least-cost method of adjustment. Apart from a weak attempt to respond to the American challenge by a Community industrial policy in advanced technology, there was no call on substitutes. Quite the contrary, a liberal common commercial policy in manufactured goods was complementary in giving even more sway to the method of market adjustment.

Industrial policy, here defined for simplicity as a set of non-mandatory government incentives such as aid and discriminatory procurement to influence structural industrial change, has the effect of altering the speeds of various sectoral adjustments. Although there exists a number of strictly economic justifications for specific forms of industrial policy – they relate to the assurance of the adequate functioning of market adjustment – the practice of the 1970s[24] in the European Community moved away from such approaches in assuming more and more conservative overtones. The Commission had to concentrate on rescuing the customs union, and with it the notion of undistorted competition throughout the Community, by pursuing a trial-and-error strategy of surveillance and prohibition of national aid schemes on the basis of Articles 85/86 and 92 to 94 of the EEC Treaty. Apart from the special case of shipbuilding, the sectors recurring are synthetic and natural textiles, clothing, footwear and steel. This industrial policy is quite an imperfect substitute for market processes of adjustment. National aid schemes tend to slow down, if not block, adjustment. The Community, its formal emphasis on the necessity of adjustment[25] notwithstanding, actually appears to get no further than the harmonization of the national degrees of slowing down adjustment.

The costs of such an industrial policy tend to be higher than in case of

market adjustment for two reasons. First, the cumulative costs over time tend to be higher since relatively low-productivity sectors can remain longer in business or can afford to improve efficiency and performance later. Secondly, there is a penalty to efficient competitors, be they domestic or foreign, because their market shares will grow more slowly than otherwise. Aid to ailing producers therefore tends to undermine the incentive system for the maintenance of permanent efficiency and for new entrants. Also, the allocation of costs is at least partly shifted to non-adjusters like the rest of the domestic economy (via the budget) and to foreign suppliers. In defence of industrial policy it should be said that at least the domestic costs of subsidies are visible for politicians and the public (and so perhaps subject to political pressures to minimize aid) and that price competition remains (so consumers do not suffer). Of course, other instruments such as government procurement from weak domestic producers do have concealed costs.

Trade policy can be effective in influencing the speed of adjustment of comparative (dis)advantage products. Within the EC the customs union implies the transfer of domestic jurisdiction on trade policy to common decision-making. Thus, trade policy did not serve as an alternative for market adjustment in the early 1960s when market integration was in full process. But it can play a role as an alternative to market adjustment to world competition, or of industrial policy. The Common Market increasingly turns to conservative trade policy as a "superior" option to either two. As Olechowski and Sampson (op. cit.) show, such protection is heavily biased towards less-developed countries which, in manufactured products, often means the NICs. In fact the Community's common commercial policy (CCP) has a Janus face in that it benignly pursues the improvement of access for non-socialist LDCs (Lomé Conventions, ASEAN, Maghreb, Generalized System of Preferences) while using plain commercial power to extract export restraints as soon as their supplies call forth resistance to adjustment.

The costs of such a trade policy are higher than in the case of industrial policy but vary with the nature of the protectionist instrument.[26] If we concentrate on "voluntary export restraints" (VER) imposed upon NICs, the costs will be higher for the domestic consumers (and this will outweigh the costs otherwise borne as taxpayers). Also, the penalty for NIC producers is more severe as efficiency and performance cannot possibly improve their access, being quantitatively given. Since VERs are discriminatory, the victimized NICs observe their competitors penetrate the same markets, which deals another blow to the incentive system they believe they can usefully employ for development. If they then diversify successfully (with all the costs of adjustment) they may run into a new dead-end street. Finally, the costs of maintaining low-productivity sectors in advanced countries tends to be higher under VERs because market shares are even

less likely to decrease than under industrial policy unless VERs apply to very few out of a multitude of suppliers and permit import growth. However, if VERs quickly spread to all major suppliers and become stricter all the time, trade policy ceases to be an instrument of adjustment. Rather it becomes an instrument which shelters the unwillingness to adjust to NICs.

The political economy of adjustment to NICs in the EC is now easy to sketch in its contours. At the Member State level, pressure groups that call for relief from the pains of adjustment cannot be helped through trade policy as the instrument is transferred to Brussels. Within limits, however, industrial policy is possible. The initially liberal trade policy of the EC meanwhile continues to create incentives for a series of NICs to export so that adjustment pressures increase in the sectors first aided and spread to other ones. Gradually, industrial policy becomes visibly expensive and is also likely to become incompatible with the customs union rules. Since the Commission can only forbid national aid but not replace it by Community aid, pressures shift eventually to the instruments the Commission *can* use: those of trade policy. The advantages to the various decision-makers are substantial. The lobbies and pressure groups prefer the use of trade policy, especially the quantitative element of it, as it gives more certainty about the adjustment pressure to expect. The national politicians will be content because it will spare them the critique of noisy pressure groups – loudly attributing sectoral unemployment to the NICs – while relieving heavy budgetary pressures. The Commission, always eager to boost Community instruments, is in a bind, since it is independent enough to perceive the grave dangers of protectionism. Its accommodative reaction is often to go ahead with protection but soften the costs by allowing for (say) 6% import growth, by minimizing the number of victim NICs and by permitting loopholes in the text.[27] This clever game of damage limitation pays off in commercial diplomacy *and* in its relation to the Council.

But the recessionary circumstances of the late 1970s have undermined the accommodative CCP. The Commission's lax and imperfect control of imports under the first Multifibre Arrangement drew heavy criticism. Member States became more aggressive in requests, despite their trade pledge in the OECD. In 1977 France blackmailed the EC by installing unilateral quotas on several textile and clothing imports. In order to meet the threat to the customs union, the Commission had to follow suit. The common commercial policy was bound to become protectionist towards NICs in order to rescue internal market integration. Nevertheless, quotas under the second Multifibre Arrangement were subdivided and assigned to individual Member States, degrading the EC to something of a free-trade area with renewed emphasis on certificates of origin at intra-EC borders.

The political economy of the present CCP is fraught with costly imperfections. The paramount failure is that the costs of slow adjustment

are first significantly enlarged and subsequently *shifted to "non-adjusters" that have no say in the decision-making process*! They include consumers and efficient firms in the EC using imported inputs from NICs that are controlled, but above all they include the workers in NICs (their management too, depending on "rents" and possibilities for export collusion). On the Community side it ought to be said that commercial policy-making is dangerously hidden from the public and is out of control for national parliaments and the European one. Lobbies talk directly to Eurocrats and sometimes the latter lean on the expertise of the former in international fora or negotiations. Until now the process has not slipped out of hand completely due to the interest some member countries still have in exports within the category of "other manufactured goods", including footwear, travel goods and handbags (Italy) and textiles and clothing (Germany, Italy).

On the NICs side, it is increasingly realized that bargains are becoming impossible. One can defend consultations between NICs and the EC about development plans that focus on key export products in need of market access to the Community (Hager, 1980), avoiding "rifle-shot" incidents. One could even argue that a selective safeguard clause in the GATT to deal with market disruption is superior to its present Article 19 (that is non-discriminatory), if only it were temporary in application and under multilateral surveillance, ensuring minimal costs to NICs. But it is a sham to compel NICs to "agree" to VERs that are unilateral quotas in all but name.

It is the good fortune of the NICs as a group that present trade policy in the EC is little more than a "fire brigade" activity.[28] Time is fully consumed by extinguishing fires in commercial diplomacy or in intra-EC bargains on the nitty-gritty product level. As we have seen in Section 4, policy being discriminatory and product-by-product, the dynamism of young NICs in textiles and clothing, and of all NICs in products where protection does not yet cover the "up-market" goods, still beats the CCP. Where it does not, it is the recessionary demand (synthetic fibres, certain steel semi-manufactures) rather than the CCP which seems to be the primary cause.

## 6. Improving the Community's relations with NICs

The European Community has substantial interests in rapidly improving commercial relations with newly industrializing countries. The EC ought to develop a long-run commercial policy *vis-à-vis* the NICs based on a profound investigation of the international and domestic, political and economic costs and benefits of alternative strategies. The case for a liberal

and adaptive policy embedded in a well-developed diplomatic framework of conflict management is a strong one. It rests on two sets of interests for the Community, still apart from the interests of the NICs.

The first set of arguments has to do with the costs of the current policy, sliding ever more into a plain servant of low-productivity, high-costs producers' lobbies. Not only do the diplomatic and political frictions caused by shifting the costs of the unwillingness to adjust onto relatively poor trading partners lead to harmful attitudes and stiffened NIC protection, the Common Market bears heavy domestic costs as well. These costs include the static losses from misallocations, the losses of efficiency in sheltered firms and the elusive, but ultimately crucial loss of dynamism as the incentive system for efficiency and performance is consistently penalized by aid schemes and protection to non-performers. In an interesting study (EC, 1979, Chapter III, 5) it is shown that the Community has a fair record up to the mid-1970s in adjusting to global structural change. The Japanese example is superior, however, and underscores the importance of anticipating and facilitating adjustment rather than accommodating low-productivity sectors.

An obstacle for a market-oriented Community adjustment policy is the separation between industrial and commercial policy. There seems to be no sensible economic reason why "Internal market" and "Industrial policy" are resorting under one Directorate-General whilst trade policy, being a partial substitute for the latter, has no organic link with it. Even the mere negative competences of the Commission in industrial policy could be applied with explicit reference to adjustment clauses in trade policy agreements. Toughness in judging aid schemes does not square with shambles in trade measures in (often) exactly the same sectors.

The second set of arguments has to do with commercial power. Commercial power *vis-à-vis* the NICs can no longer be taken for granted. The EC has been lulled by the ease with which it played the bilateral, sectoral power game after generally agreeing on the second Multifibre Arrangement in early 1978. Basic to the commercial power relation between NICs and the EC is the asymmetrical importance of mutual trade to overall domestic economic activity for the two parties. As Table 3 and other data clearly suggest, access to the EC is much more important to the NICs than the NIC trade is for the EC. There still is not one single sector, as defined in Table 4, in which the share of no less than 16 NICs is larger than the share of intra-EC imports in total imports (although some sectors in 1979 came rather close). The share of NIC imports in EC production would be even lower.[29]

But other power factors have changed in the late 1970s. Though the big surplus of the EC in manufactures trade with NICs is relatively decreasing, this is predominantly due to Spain, South Korea and Taiwan. With Hong Kong the EC ran a deficit in 1979 but recent growth rates of EC

exports to Hong Kong were much higher than those of imports. Other NICs provide the Community with a comfortable surplus, that enhances the Community's vulnerability in times of heavy overall deficits. NICs could start exploiting this.

More important, however, is the stake in the volume EC-NIC trade has now reached. Is it realized in Brussels that EC 1979 imports from (16) NICs in "machinery and transport equipment" and in "other manufactured goods" amounted to 93% of those from the USA, up from 68% in 1975? Has one observed that the 1979 *exports* of "machinery and transport equipment" of the Common Market to 16 NICs are one-third *higher* than those to the USA? Is it known that EC 1979 exports in "other manufactured goods" to the USA and to the NICs are at par while EC chemicals exports to NICs are nearly one and a half times higher than those to the USA? Even if Japan and the USA are combined, EC 1979 exports in chemicals and in "machinery and transport equipment" to NICs are still equal or larger than to these big traders.

These stakes in trade are not in the least reflected in the conduct of commercial policy. With the USA and Japan complex bilateral networks of information and regular consultation have been built up, including the highest political level in world economic summits, and incorporating multitudinous contacts in multilateral fora such as GATT and the OECD (and its Steel Committee, for instance). For NIC-EC trade such a framework is lacking altogether. Contacts are occasional and the style of diplomacy reminds one of old-fashioned tariff wars. The EC may even be forced into more regular consultation and bargaining on equal footing if the NICs were to found a common organization. Such an "NIC caucus" would not be easy to set up but the formation of an effective coalition of Lomé associates has shown that it is a realistic possibility. A tightly managed NIC caucus would have several powerful weapons. It could bring in more authority by blocking the Community's bilateral approach to NICs. It could also bargain over a wider scale of products than any single NIC could do. Its ultimate weapon would be a buyer's boycott on selected products in view of the sizeable EC surplus. A more subtle route could consist in discriminatory retaliation with simultaneous offers for trade liberalization in response to blatant forms of EC protectionism. Since the NICs, except Singapore and Hong Kong, are highly protectionist themselves, there is ample room for organizing countervailing power. Finally, an NIC caucus could facilitate prior consultation about development plans with a view of preventing "rifle-shot" incidents.

It is highly desirable that a more adequate management of EC-NIC trade relations be established, reflecting an awareness of the costs of the recent inward-looking commercial policy and of the substantial and growing stakes the Community has in trading with newly industrializing countries.

## References

B. Balassa, "The changing pattern of comparative advantage in manufactured goods", *Review of Economics and Statistics*, Vol. LXI, 2, May 1979.
V. Cable, *World textile trade and protection*, Economist Intelligence Unit Special Report No 63, London, 1979.
EC, "Changes in industrial structure in the European economies since the oil crisis, 1973-78", *European Economy*, special issue, October 1979.
A. Edwards, *The new industrial countries and their impact on Western manufacturing*, Economist Intelligence Unit Special Report No 73, 2 vols, London, December 1979.
W. Hager, "Responding to the challenge", in F. Duchêne (ed.), *New and old industrial countries in the 1980s*, 1980 (forthcoming).
D. Keesing and M. Wolf, *Textile quotas against developing countries*, Trade Policy Research Centre, Thames Essay No 23, London, 1980.
D. Lal, *Market access for semi-manufactures from developing countries*, Sijthoff, for the Trade Policy Research Centre and the Graduate Institute of International Studies, Leiden, 1979.
T. Murray, W. Schmidt and I. Walter, "Alternative forms of protection against market disruption", *Kyklos*, Vol. 31, 4, 1978.
D. Nayyar, "Transnational corporations and manufactured exports from poor countries", *Economic Journal*, Vol. 88, March 1978.
B. Nowzad, 1978, *The rise in protectionism*, IMF, Pamphlet series No 24, Washington DC, 1978.
A. Olechowski and G. Sampson, "Current trade restrictions in the EEC, the United States and Japan", *Journal of World Trade Law*, Vol. 14, 3, May/June 1980.
OECD, *L'incidence des nouveaux pays industriels sur la production et les échanges des produits manufacturés*, Paris, 1979.
S. Page, "The increased use of trade controls by the industrial countries", *Intereconomics*, Vol. 15, 3, May/June 1980.
J. Pelkmans, "How effective is Western economic cooperation?", *Auçenwirtschaft*, Vol. 35, II, 1980.
L. Turner, C. Bradford, L. Franko, N. McMullen and S. Woodcock, *Living with the newly industrializing countries*, Royal Institute of International Affairs, London, Chatham House Papers No 7, 1980.
Unctad, *Handbook of international trade and development statistics*, New York, 1979.
World Bank (IBRD), *World development report*, Washington DC, 1979.
World Bank (IBRD), *World development report*, Washington DC, 1980.

## Notes

[1] Such countries and territories as Ecuador (27.4%), Trinidad and Tobago (22.6%) St Lucia (24.2%), Bahamas (43%; the record), Cuba (23.2%), Tunisia (22.9%), Congo (27.0%), Guinea (26.9%), Rwanda (22.4%), Botswana (27.6%), Brunei (41.6%) and Macao (23.9%). The list in incomplete. All rates over 1970-78.

[2] Venezuela falls out as well as it is an OPEC member. In the near future ignoring it might be incorrect. From 1975-79 EC manufactures imports from Venezuela soared from 4 million (EUR) to 77 million ECU. This absolute value is however still insignificant on the big Common Market. (For EUR and ECU, see below, Table 2).

[3] *Import penetration* is defined as imports (here, from NICs) as a percentage of apparent domestic consumption. Apparent consumption is (domestic) production plus imports,

minus exports. Import penetration has to be distinguished from *import share*: imports, of a certain product or from a certain country, as a percentage of total imports. The two yardsticks may give very different pictures! See also Section 3.

[4] The expression is in Turner et al. (1980, p. 20). The celebrated example is about three South Korean plants producing TV sets for the US market (only!) and achieving an output of some 400 000 units a year each. US protection forced them down to half. Here, scale economies do cause market disruption.

[5] One might have doubts whether EC access to the Japanese market is indeed comparable to Japanese access to the EC. Surely, there is evidence that hidden barriers — be they hidden in the distribution system, or, in the allocation of television time, or otherwise — tend to suppress the inflow of EC products into Japan. But there is also evidence that EC firms deliberately avoid the problems of unacquaintedness with the Japanese market by opting for third markets. Futhermore, EC members have a poor record of providing several quotas against Japan for nearly three decades. Most NICs are so protectionistic that market access is plainly not comparable with the access of NICs to the EC. The more developed they are the less appropriate this is. In brief, if an NIC "graduates" in terms of value and composition of trade, it should also graduate with respect to the basic principles of Gatt, rather than remaining exempted on the basis of a status it is overcoming (to wit: being poor).

[6] See footnote 1, Table 2 for EUR and ECU.

[7] Note that South East Asian NICs have recently increased their chemicals imports from the EC at conspicuously high rates. This might have to do sith the synthetic fibre industry in these NICs.
It should also be of interest to know that the 1979 EC surplus in chemicals trade with NICs is more than five times (!) the absolute value of imports from NICs.

[8] Portugal has a low export level (to the EC) of these goods, but a considerable export base. Its GNP per capita is also much lower.

[9] Ireland is left out. As discussed before, chemicals will not be considered since there seems to be hardly any adjustment problem. For a comparison with Table 3, note that the NIC share in EC chemicals imports was 2.2% both in 1975 and in 1979.

[10] Except for Denmark.

[11] The most important reason is that trade and industry statistics are insufficiently harmonized as to their nomenclatures. Penetration rates, if published at all, are usually several years behind, and rather aggregate.

[12] The SITC used here is the first revision. A second revision has caused further changes, but does not touch upon the nature of the argument of course.

[13] In addition, high degrees of disaggregation can only be used selectively as one runs into very serious data problems. The OECD, for instance, is not consistent over the years in presenting the (same) five-digit product groups or in presenting a country breakdown for them.

[14] In "wood manufactures, not elsewhere specified" (SITC 632; boxes, cases, crates, builders' woodwork, etc.) there semms to be little if any such problem. This has to do partly with the preponderance of the raw material input and partly with the already achieved specialization in quality wood in many EC countries. Beyond any doubt is the case of cork manufactures (SITC 633) where Portugal and Spain together reach a share of 83.4% of EC imports in 1975 and a Community industry hardly exists. Furthermore, TV receivers is a special case (see text).

[15] This is higher than the actual weighted average. The consumer price deflator and the import price deflator for the EC are respectively: 1976 (10.1 and 11.5%); 1977 (9.7 and 8.3%); 1978 (7.1 and 0.5%) and 1979 (9 and 10%). *European Economy*, March 1980, p. 129.

[16] See Turner et al. (1980), p. 8) where importance of direct investments is argued to be minor. For a survey, see Nayyar (1978).

[17] Edwards (1979, Vol. II p. 32) for a similar "suspicion".

[18] MFN triffs are tariffs for imported goods from all countries that are contracting parties to Gatt and therefore enjoy most-favoured-nation treatment.

[19] It should not be forgotten that this ignores so-called "effective protection": the fact that tariffs are higher the higher the stage of production. It is known that the structure of effective protection (also) in the EC is significantly biased against processing of raw materials and production of semi-manufactures in LDCs. See Lal (1979) for an extensive empirical survey.

[20] See especially Nowzad (1978), Murray, Schmidt and Walter (1978) and Turner et al. (1980, p. 15). The following examples are taken from Nowzad (1978, App.s. X, XIII and XIV; manufactures only). At the end of 1977 Taiwan suffered from an import prohibition (!) in radios (France), a licence scheme in umbrellas (France) and a quota for umbrellas (Germany). In the 1970s South Korea suffered from a licence scheme in fottwear (United Kingdom) and quota for black and white TV sets (United Kingdom), for radios (France; two temporary ones), for umbrellas (France; two temporary ones), for silk fabrics and for toys (France), as well as licence schemes in cassette recorders and cutlery (Denmark). At the end of 1977 the Philippines suffered from a tariff quota in veneers and plywood (EEC; initial zero tariff). The list does not include protection under the Multifibre Arrangement.

[21] Here, LDCs include the Middle East. Note that she applies 1974 trade figures to 1979 restrictions in order to avoid that the import value of controlled items would shrink *due to such controls*.

[22] For excellent treatises, see Keesing and Wolf (1980) and Cable (1979).

[23] See OECD (1979, Annex 2) for an extensive survey.

[24] Apart from steadily weakening French indicative planning and italian interventionism, industrial policy in the 1960s was mostly paper work, both at the EC and country level. The conspicuous moves by countries (Concorde, aluminium, gas, oil) were primarily related to economic security. Regional industrial policy also played a role.

[25] By such norms as: aid ought to be temporary, ought not have capacity-enlarging effects, etc.

[26] For a distinction between hard and soft protection on the basis of the level and allocation of the costs of protection, see Pelkmans (1980).

[27] Threatened NIC producers can also be lured into restraints by allowing them to seize the rent. This can be done by giving rather than selling established producers the import licences.

[28] Expression of an EC trade policy official in private discussion with the author.

[29] See, for some exples of disaggregated sectoral import/production rations, Nowzad (1978, p. 105). The 1976 rations for French imports from all LDCs never come above 16%. For import penetration rates (see footnote 3) of some clusters of two-digit sectors, for the EC, in 1975, see Unctad, 1979, Table 7.1. For manufactures from all LDCs they never reach beyond the 7% of clothing.

# Heterogene Erwartungen, Preisbildung und Informationseffizienz auf spekulativen Märkten

MANFRED E. STREIT*
Mannheim/Florenz

*A disquieting feature of aggregate models which assume uniform expectations is that they don't explain why there are any transactions at all in existing assets.*
*(James Tobein (1980))*

Ausgehend von der Voraussetzung heterogener Erwartungen der Marktteilnehmer sollen im folgenden (1) der Preisbildungsprozeß auf Terminkontraktmärkten erläutert und (2) die Konsequenzen dieser Betrachtungsweise für die Informationseffizienz solcher Märkte geprüft werden. Terminkontraktmärkte als organisierte Form der allgegenwärtigen spekulativen Märkte wurden gewählt, weil ihre institutionellen Gegebenheiten klare Bezugspunkte darstellen[1]. Diese wiederum erleichtern es, die theoretischen Vorstellungen zu verdeutlichen.

## Ein fast vergessenes Argument

Die Bedeutung heterogener Erwartungen für die Preisbildung auf Termin- bzw. Terminkontraktmärkten wurde zwar früh erkannt, aber auf eine Weise verdrängt, die auch für jüngere Modellierungsversuche gilt. Dies wird besonders deutlich an einer Diskussion zwischen Kaldor und Hawtrey vor mehr als vierzig Jahren[2]. Kaldor räumte auf einen entsprechenden Einwand Hawtrey's zwar ein, daß in seiner vorangegangenen Analyse[3] die Preisbildung auf diesen Märkten inadäquat dargestellt worden sei, da er die Erwartungsvielfalt vernachlässigt habe. Zugleich machte er jedoch geltend, daß eine Argumentation mit "repräsentativen" Erwartungen es erlaube, das Problem zu vereinfachen, ohne das Ergebnis materiell zu beeinflussen. Es gäbe immer eine Erwartung die zum tatsächlichen Preis führe, wenn sie von allen Marktteilnehmern gehalten würde[4]. Das Argu-

ment Hawtrey's, daß Spekulation mit homogenen Erwartungen nicht erklärt werden könne[5], wurde fast vergessen. Die Vereinfachung schien auch analytisch unproblematisch zu sein, erlaubte sie es doch, die Präsenz von Spekulanten auf Keynes'sche Weise zu erklären, nämlich daß ihnen von den Hedgern eine Prämie für die Übernahme des Preisänderungsrisikos gezahlt werde.

In einem Großteil der sich anschließenden theoretischen Diskussion wurden zwei empirische Befunde übersehen: (1) Auch mit verbesserten statistischen Methoden war der Nachweis einer hinreichend weitverbreiteten, systematischen Verzerrung von Termin- bzw. Terminkontraktpreisen nicht zu führen; eine solche Verzerrung ist jedoch als Beweis für die Existenz einer Risikoprämie zu fordern[6]. (2) Das beobachtbare Verhalten von Hedgern paßt nicht zum Modellverhalten; der Anteil diskretionären Hedgings ist zu beträchtlich, um an der Vorstellung festhalten zu können, Hedger ließen sich nicht von eigenen Preiserwartungen leiten, sondern trachteten routinemäßig danach, ihr Preisänderungsrisiko abzuwälzen.

Wie attraktiv die Vorstellung "repräsentativer" Erwartungen unter Modellierungsaspekten geworden ist, zeigte sich auch in der jüngsten Zeit. Vor allem bei dem Versuch, Informationsaspekte des Handels auf spekulativen Märkten zu modellieren, werden die "repräsentativen" Erwartungen vorherrschend als rationale Erwartungen konkretisiert. Sie zu bilden vermögen in erster Linie Spekulanten mit Hilfe der von ihnen erworbenen Informationen. Dies bedeutet zwar eine Akzentverschiebung weg von der schlichten Risikoübernahme, jedoch ergeben sich Schwierigkeiten bei dem Versuch, Informationsbeschaffung und -verwertung in spekulativer Absicht, Informationseffizienz des Gleichgewichtspreises sowie Informationsübertragung durch Preissignale mittels rationaler Erwartungen und i.d.R. im Rahmen einer komparativ statischen Analyse überzeugend abzubilden[7]. Vor allem ist es unter den Modellbedingungen schwierig, Anreize zur Informationsbeschaffung zu begründen; denn der einzelne Marktteilnehmer vermag Prognosevorteile aus dem Informationserwerb für sich nicht ohne weiteres zu internalisieren, da Nichtinformierte jede in den Gleichgewichtspreis eingeflossene Information identifizieren können. Nur wenn die Informationsübertragung über den Preis durch einen stochastischen Prozeß gestört wird, gelingt es, eine Informationsrente zu erzeugen, die auf informierte Spekulanten verteilbar ist[8]. Ohne solche Störungen bleibt in derartigen Modellen zwar der Rückgriff auf unterschiedliche Risikoneigungen als Erklärungsmöglichkeit für Handel zwischen letztlich Gleichinformierten, aber die Anreizproblematik im Hinblick auf Informationsaktivitäten stellt sich erneut[9]. Generell dürfte auch für die Mehrzahl der jüngeren modelltheoretischen Versuche, das Geschehen auf spekulativen Märkten abzubilden, der grundsätzliche Befund KIRZNERS ((1981), S. 121) zutreffen, daß der hohe Preis i.S.v. Verlust an fundamentalem ökonomischem Verständnis ignoriert wird, der für die Anwendung hoch-

entwickelter technischer Methoden von zweifelhaftem praktischem Wert zu zahlen ist.

Im folgenden wird argumentiert, daß beides, Bemühungen, Preisänderungsrisiken abzuwälzen, und Versuche, prognoseverbessernde Informationen aufzuspüren und zu verwerten, zentrale und engverknüpfte Aspekte des Terminkontrakthandels sind. Der Preisbildungsprozeß dürfte sich weder adäquat dadurch erklären lassen, daß nur auf einen der Aspekte abgestellt wird, noch dadurch, daß die Marktteilnehmer in zwei entsprechende Gruppen – risikoscheue Uninformierte und weniger risikoscheue Informierte – aufgespalten werden. Darüber hinaus dürfte das Informationsproblem nur mit einer dynamischen Analyse adäquat erfaßbar sein.

## Risikoneigung und Informationsaktivität

Eine enge Verknüpfung von Risiko- und Informationsaspekten dürfte für das selektive Hedging charakteristisch sein. Der Informationsaspekt läßt sich erschließen, wenn berücksichtigt wird, unter welchen Bedingungen zum Beispiel ein Short-Hedging sich als vorteilhaft erweist, verglichen mit einem Verzicht auf Absicherung der gehaltenen oder erwarteten Kassaposition, also einer de-facto-Spekulation auf dem Kassamarkt. Dies ist der Fall, wenn sich die Differenz zwischen Terminkontrakt- und Kassapreis (die Basis) zugunsten des Hedgers verändert[10], was stets eine Änderung des Terminkontraktpreises erfordert. Letzteres gilt auch für den Grenzfall des perfekten Hedgings (konstante Basis). Hingegen entstehen bei einem konstanten Terminkontraktpreis lediglich zusätzliche Transaktionskosten[11].

Was bedeuten die vorgenannten Bedingungen zunächst hinsichtlich der Erwartungen, die bei selektivem Hedging unterstellt werden können? Wenn die Marktteilnahme eines selektiven Hedgers[12] gezielt und erfolgversprechend sein soll, kann sie nicht auf der Einschätzung beruhen, der herrschende Kontraktpreis stelle de facto eine zutreffende Prognose des Kontraktpreises bei Kontraktfälligkeit dar[13] denn dann wären aus der Sicht des potentiellen Hedgers nur vermeidbare Transaktionskosten zu erwarten. Vielmehr ist die Einnahme einer Kontraktmarktposition nur interessant, wenn der potentielle Hedger eine von der de-facto-Prognose des Terminkontraktpreises abweichende Erwartung hat und ihr vertraut. Seine Erwartung kann beruhen (1) auf Intuition, aber auch und vor allem (2) auf spezifischen Informationslagen. Letztere wiederum können eine abweichende Beurteilung allgemein verfügbarer zukunftsträchtiger Informationen beinhalten oder die Auswertung von Informationen, die anderen Marktteilnehmern noch nicht zugänglich sind.

Für eine Verknüpfung von Risiko- und Informationsaspekten ergibt sich, daß die Chance, das Preisänderungsrisiko mittels Hedging abwälzen

zu können, durch Informationsaktivitäten verbesserbar sein dürfte. Zumindest kann ein Hedger die Informationen heranziehen, die er mit seiner Tätigkeit als Produzent, Händler oder Verarbeiter der Kassaware erschließt. Demgegenüber würde der Routinehedger entweder keinerlei Preiserwartungen bilden oder aber ihnen niemals vertrauen[14]. Er würde sich stets damit begnügen, einen Terminpreis als Kalkulationsgrundlage zu fixieren.

Für die Preisbildung ist von Belang, daß das, was hinsichtlich der Erwartungen eines selektiven Hedgers vermutet wird, erst recht für Spekulanten im konventionellen Sinne gelten dürfte. Verallgemeinernd und einige Einschränkungen vorläufig zurückstellend läßt sich vermuten: Für solche Marktteilnehmer ist eine Terminkontraktposition nur dann erwägenswert, wenn sie zu einem Preis eingenommen werden kann, der aus ihrer Sicht – ihrer Intuition und Informationslage – eine falsche de-facto-Prognose des Preises bei Kontraktfälligkeit darstellt. Da dies für beide Seiten einer Transaktion gilt, dürfte auf Kontraktmärkten Handel zu "falschen" Preisen eher die Regel denn die Ausnahme sein.

Um unter diesen Umständen die Preisentwicklung eines Kontraktes bis zu seiner Fälligkeit darstellen zu können, ist es erforderlich, die zeitliche Struktur der Transaktionen zu berücksichtigen. Zwei Zeitpunkte sind jeweils von Belang: der Zeitpunkt, in dem eine Kontraktmarktposition eingenommen wird, und der, in dem diese Position – i.d.R. durch Einnahme einer entsprechenden Gegenposition – wieder aufgehoben wird. Aus der Perspektive das Handels zu "falschen" Preisen ist entscheidend, daß eine Kontraktmarktposition nur in der Erwartung eingenommen wird, sie später gewinnbringend aufheben zu können[15]. Hierin liegt auch das spekulative Element des (diskretionären) Hedgings[16].

Nunmehr ist zu prüfen, unter welchen Umständen eine zum "falschen" Preis eingenommene Terminkontraktposition mit Gewinn aufgehoben werden kann. Dies ist der Fall, wenn die den zukünftigen Kontraktpreis umgebende Ungewißheit ganz oder zum Teil in einer Weise beseitigt wird, die konsistent ist mit der Intuition bzw. Informationslage, auf der die Erwartungen eines Marktteilnehmers beruhen. Völlig beseitigt ist die Ungewißheit bei Fälligkeit eines Kontraktes; denn zu diesem Zeitpunkt diktiert die tatsächliche Lage auf dem Kassamarkt (über die Arbitragebeziehungen) eines Konsens auf dem Kontraktmarkt.

Teilweise beseitigt wird die Ungewißheit[17] in der Periode vor Fälligkeit eines Kontraktes in einer für einen Marktteilnehmer vorteilhaften Weise, (1) wenn sich die Informationslage anderer Marktteilnehmer der seinen anpaßt und diese gleiche oder ähnliche Folgen für den zukünftigen Kontraktpreis daraus ableiten, sowie (2) wenn Informationen den Markt erreichen, die zwar von den seinen abweichen, aber dennoch eine Preiserwartung begünstigen, die der seinen ähnlich ist. Unter diesen Umständen würden die geänderten Erwartungen genügend Marktgewicht haben, um

eine Kontraktpreisänderung zu bewirken, die mit dem nunmehr erwarteten Preis vereinbar ist. Die Situation spiegelt einen partiellen Konsens der neu informierten, möglicherweise aller Marktteilnehmer wider. Er bezieht sich auf die Richtung, in der der herrschende Kontraktpreis zu korrigieren ist. Um es jedoch dem einzelnen Marktteilnehmer zu ermöglichen, aus seiner Sicht erfolgversprechende Positionen einzunehmen, muß er wiederum zu einem Kontraktpreis abschließen können, der seiner Informationslage bzw. den daran anknüpfenden Erwartungen nicht vollständig entspricht. Anderenfalls würde sich ein Engagement nicht lohnen, da er lediglich vermeidbare Transaktionskosten erwarten könnte. Das gleiche muß für diejenigen gelten, die die Einnahme einer Gegenposition erwägen. An ihnen ist es, durch ihr Engagement eine Widerstandslinie für die einsetzende Preisbewegung entstehen zu lassen. Aus dieser Sicht ergibt sich, daß in der Periode vor Fälligkeit eines Kontraktes Preisbewegungen nicht bei einem Konsens zum Stillstand kommen, der sich über die Richtung hinaus auch noch auf das Ausmaß der Korrektur der de-facto-Prognose durch den Kontraktpreis bezieht. Handel zu "falschen" Preisen ist die Regel. Jeder Preis, zu dem Umsätze in dieser Periode beobachtet werden, stellt lediglich ein transitorisches Gleichgewicht heterogener Erwartungen dar[18].

Wie angekündigt, sind die vorangegangenen Überlegungen noch etwas zu qualifizieren. Sie bezogen sich in erster Linie auf die Eröffnung neuer Positionen. Diese mögen sowohl von selektiven Hedgern als auch von Spekulanten im konventionellen Sinn eingenommen werden. Darüber hinaus mag eine Preisbewegung die Preisstruktur der verschiedenen gehandelten Kontrakte ebenso wie die Basis verändert haben. Als Folge können lohnende Spreading- und Arbitragemöglichkeiten entstanden sein. Auch ihre Wahrnehmung beinhaltet aber i.d.R. die Erwartung, daß sich der Kontraktpreis ändert[19].

Darüber hinaus gibt es jedoch noch drei weitere Gründe für Transaktionen, bei denen die Erwartungen nicht unbedingt von der de-facto-Prognose des Kontraktpreises abweichen müssen, zu dem kontrahiert wird. Ein Teil der Transaktionsnachfrage, die Preisänderungen auslöst, mag zur Einnahme von Gegenpositionen bei denjenigen anregen, denen es gelang, eine durch ihre Informationslage begründete Position einzunehmen, bevor ihre Erwartungen breitere Bestätigung fanden. Sie finden sich bestätigt und schreiten nun zu Gewinnmitnahmen[20] Ein weiterer Teil mag von Marktteilnehmern kommen, die Verluste aus zuvor eingenommenen Positionen begrenzen möchten. Sie würden im Einklang mit der beobachteten Preisbewegung eine Revision ihrer ursprünglichen Erwartungen vornehmen. Schließlich kann es sein, daß einige Marktteilnehmer, vor allem sog. Scalper, neue Positionen eingehen, obgleich sie keine vom herrschenden Preis abweichenden Erwartungen haben. Auf diese Weise mögen sie sonst nicht akkomodierbare Transaktionswünsche einschließlich

(netto) Routinehedging kurzfristig befriedigen. Der Preis dafü, daß sie "einen Preis machen", kann in Kommissionen und einem größeren Geld-Brief-Unterschied gesehen werden[21]. Letzteres ist aber durchaus wieder mit Hypothese des Handelns zu "falschen" Preisen verträglich; denn eine vorübergehende Vergrößerung des Geld-Brief-Unterschieds[22] erlaubt es dem Scalper, von dem erwarteten Preis abzuweichen und eine ausreichende Gewinnchance zu schaffen. Für den Routinehedger würde daraus die Prämie erwachsen, die er für die Abwälzung des Preisänderungsrisikos zu zahlen bereit ist.

## Temporäre Konvergenz heterogener Erwartungen

Bleibt zu erläutern, wie es zu der beobachtbaren, mit Annäherung an den Fälligkeitstermin eines Kontraktes sich verbessernden Prognosequalität des Kontraktpreises kommt. Allgemein läßt sich vermuten, daß mit der Annäherung an den Fälligkeitstermin die Informationsdichte über diesen zukünftigen Zeitpunkt zunimmt. Dazu gehört auch, daß sich i.d.R. die Zahl der grundsätzlich möglichen Ereignisse, die den Preis bei Fälligkeit beeinflussen können, mit größerem Vertrauen verringern läßt. Ferner dürften generell die Transaktionsbeziehungen zwischen Gegenwart und naher Zukunft häufiger sein als mit der fernen Zukunft.

Der arbeitsteilige Prozeß, mit dem die sich verdichtenden Informationen in das Kontraktmarktgeschehen eingeschleust werden, bereitet besondere Modellierungsschwierigkeiten. Was nämlich abzubilden wäre, ist die Art und Weise, wie die mit Transaktionswünschen angezeigten Erwartungen und damit verknüpfte Informationslagen im Verlaufe des Marktgeschehens geprüft und u.U. verbreitet werden. Wichtig dürfte dabei sein, daß Erwerb, Prüfung und Diffusion von Informationen aus der Perspektive einer wesentlich größeren Vielfalt von Transaktionen erfolgt, als sie normalerweise modelliert wird[23].

Dabei beziehen sich die Informationen bzw. die von ihnen beeinflußten Erwartungen:

– auf einzelne Kontraktpreise und deren Veränderung im Falle der Spekulation konventioneller Art;

– auf Differenzen zwischen Kontraktpreisen und den entsprechenden Kassapreisen sowie darauf, wie es vermutlich zu Veränderungen dieser Differenzen kommt, d.h. die verschiedenen, unterschiedlich motivierten Formen des Hedgings;

– auf Differenzen zwischen (a) Preisen für den gleichen Kontrakt aber für unterschiedliche Fälligkeiten, (b) Preisen für verschiedene Kontrakte für substituierbare Produkte aber gleiche Fälligkeiten, (c) Preisen für verschiedene Kontrakte der gleichen Ware an unterschiedlichen Börsen-

plätzen sowie deren Veränderung, d.h. die verschiedenen Formen des Spreadings;

– auf Differenzen zwischen (a) Preisen für den gleichen Kontrakt an verschiedenen Börsenplätzen und (b) Preisen für einen Terminkontrakt und den entsprechenden Kassapreisen, d.h. die reine räumliche und zeitliche Arbitrage.

Hinzu kommt, daß sich Marktteilnehmer häufig auf eine oder wenige dieser Transaktionsarten spezialisieren[24]. Daraus ergibt sich, daß die Preisentwicklung ständig aus sehr unterschiedlichen Perspektiven auf Gewinnchancen überprüft wird, die sich nicht zuletzt aus der Einschätzung individueller Informationslagen ableiten. Die Folge entsprechender Transaktionen ist ein spontaner Prozeß des Probierens und Korrigierens, über den eine Informationsanreicherung im Kontraktpreis erzielt wird[25]. Diese erfolgt vor allem bei lagerfähigen Gütern in engem Kontakt mit der Entwicklung auf dem Kassamarkt. Informationsverdichtung und zeitliche Arbitrage dürften bewirken, daß die Heterogenität der Erwartungen mit Annäherung an den Fälligkeitstermin eines Kontraktes abnimmt und daß Kassa- und Kontraktpreis konvergieren.

Wie bereits dargelegt, kommt es bei Fälligkeit eines Kontraktes aufgrund der faktischen Situation auf dem Kassamarkt notwendig zu einem Konsens über den Kontraktpreis. Dieser Konsens ist jedoch höchst temporär und nicht frei vom (indirekten) Einfluß durchaus heterogener Erwartungen. Zweierlei ist nämlich zu berücksichtigen: Einmal ist der Kassapreis selbst ein spekulativer Preis, beeinflußt z.B. von erwartungsbezogenen Lagerhaltungsdispositionen. Zum anderen gleitet gewissermaßen die Zeit auf dem Kontraktmarkt unter einer ganzen Serie von Kontrakten mit unterschiedlichen Fälligkeiten hinweg. An sie werden entsprechend der zeitlichen Tiefe des Marktes am Ende der Fristen stets neue Kontrakte angefügt. Die einzelnen Kontraktpreise und der Kassapreis sind hochgradig interdependent aufgrund des Zusammenspiels von Spreading und zeitlicher Arbitrage. Dementsprechend macht sich die Zukunft in der flüchtigen Gegenwart des Kassa- und Kontraktmarktgeschehens auch bei Fälligkeit eines Kontraktes bemerkbar.

## Informationsaktivität und Informationseffizienz

Welche Konsequenzen ergeben sich aus dem zuvor skizzierten Prozeß für das Ausmaß an Informationssuche sowie für die Qualität der Verarbeitung und Verbreitung zukunftsrelevanter Informationen auf Terminkontraktmärkten? Die Antwort hat eine Anreiz- und eine Wettbewerbskomponente, die – wie stets bei Marktprozessen – zusammenwirken.

Die Anreizkomponente beinhaltet, daß die Marktteilnehmer eine

Chance haben müssen, Erträge aus Informationsaktivitäten zu internalisieren. Grundlage für solche Erträge ist eine Verbesserung der Treffsicherheit subjektiver Preiserwartungen als Folge von entsprechenden Informationen und ihrer richtigen Auswertung. Internalisierbarkeit eines Informationsertrags setzt voraus, daß es einem Marktteilnehmer gelingt, zu einem, verglichen mit seiner Erwartung falschen Preis zu kontrahieren. Das ist wiederum nur möglich, wenn seine Erwartung von anderen noch nicht geteilt wird, er also über einen zeitlichen Informationsvorsprung verfügt; der Vorsprung muß nicht unbedingt aus einer neuen Information bestehen, sondern kann, wie dargelegt, auch auf einer Neubeurteilung weithin verfügbarer Informationen beruhen.

Allerdings kommt ein Marktteilnehmer nicht umhin, seinen Informationsvorsprung selbst zu gefährden. Die Gefährdung ergibt sich aus seinem Bemühen, den Vorsprung durch Transaktionen zu verwerten, d.h. mit dem Versuch, eine Order zu plazieren. Der Verwertungsversuch ist mit einem externen Effekt verbunden. Dies gilt in besonderem Maße für die direkt am Kontrakthandel Beteiligten. Der externe Effekt besteht aus dem entsprechenden Preis-Mengen-Signal auf dem Kontraktmarkt. Ob, in welchem Ausmaß und wie schnell dieses Signal zu einer den dahinterstehenden Erwartungen bzw. Informationen entsprechenden Preisänderung führt, hängt davon ab, (1) inwieweit das Signal andere zu Informationsaktivitäten anregt, (2) ob einige "mit dem Markt gehen", sich also zur einfachen Information entschließen, und (3) inwieweit es zur Keynes-'schen "Schönheitskonkurrenz", also zu Versuchen anderer Marktteilnehmer kommt, das ursprüngliche Signal zu interpretieren. Hinzu kommt nicht zuletzt gewissermaßen eine Fremdgefährdung, wenn zugleich andere Marktteilnehmer durch eigene Informationsaktivitäten zu ähnlichen Erwartungen gekommen waren. Zusammengenommen beschreiben diese Handlungen die Wettbewerbskomponente.

Aus dem Zusammenwirken der beiden Komponenten ergibt sich, daß das Kontraktmarktgeschehen unter Informationsaspekten als ein durch die Suche nach zeitlichen Informationsvorsprüngen bestimmter Wettbewerbsprozeß beschrieben werden kann. Das ändert nichts an der Motivationsstruktur von Marktteilnehmern, die z.B. durch das Bemühen dominiert sein mag, Preisänderungsrisiken abzuwälzen; wie dargelegt, wird das Bemühen durch Informationsaktivitäten – allerdings nicht kostenlos – eher erleichtert. Ebensowenig sind spezifische Annahmen über die Risikoneigung erforderlich; die Risikoneigung dürfte allerdings die Entscheidung darüber beeinflussen, ob und gegebenenfalls in welcher Form aufgrund einer subjektiven Informationslage Positionen am Kontraktmarkt eingenommen, korrigiert oder aufgegeben werden[26].

Wohl aber hat diese Sicht des Kontraktmarktgeschehens Konsequenzen für die Qualität der Verarbeitung und Verbreitung zukunftsrelevanter Informationen auf Terminkontraktmärkten, also für ihre Informationseffi-

zienz. In der Periode vor Fälligkeit des Kontraktes können diese Märkte nicht effizient im Fama'schen Sinne sein, wenn der zuvor beschriebene Prozeß die Informationsaktivitäten und ihre Marktfolgen im Kern beschreibt. Dabei kommt es aus der hier vertretenen Sicht nicht so sehr auf die von Fama ohnehin konzedierte Verletzung der Bedingungen kostenloser Information und Transaktionen an[27]. Schon bedeutsamer ist die von ihm ebenfalls erwähnte Möglichkeit, daß die Marktteilnehmer die Marktfolgen einer Information unterschiedlich beurteilen. Darüber hinaus sind jedoch vor allem die Effizienzfolgen zu berücksichtigen, die sich aus dem hier betonten Handel zu "falschen" Preisen auf der Grundlage heterogener Erwartungen ergeben können:

Die Eröffnung einer Kontraktmarktposition ist überwiegend nur dann attraktiv, wenn sie zu Preisen erfolgen kann, die nach Einschätzung der an Transaktionen Beteiligten ihre individuelle Informationslage und davon abgeleitete Erwartungen nicht voll widerspiegelt. In der Chance, zu einem zunächst nur subjektiv ineffizienten Preis kontrahieren zu können, liegt der Anreiz zu Informationsaktivitäten. Ein erzielter Gewinn ist die Prämie dafür, schneller bei Erwerb und Auswertung neuer bzw. der Neubewertung verfügbarer Informationen gewesen zu sein. Die Prämie ist höchst ungewiß und stets durch den Wettbewerb anderer Marktteilnehmer gefährdet. Dieser Wettbewerb um temporäre Informationsvorsprünge bestimmt Umfang und Geschwindigkeit der Informationsverbreitung auf dem Kontraktmarkt.

Jedoch erlaubt die bisherige Argumentationsweise kaum ein abschließendes Urteil über die Informationseffizienz des Kontraktmarktes. Dazu ist der Sprung vom dem einzelnen Marktteilnehmer auf die Aggregationsebene des Marktes erforderlich. Nicht überzeugend scheint in jedem Fall die bislang modelltheoretisch dominante Praxis zu sein, im wesentlichen auf den Marshall'schen repräsentativen Marktteilnehmer und mit ihm auf homogene Erwartungen zurückzugreifen. Auf diese Weise wird zwar ein Gleichgewicht modellierbar, jedoch nur unter praktischem Verzicht auf das spekulative Element.

Bei dem hier eingeschlagenen Weg ist die Schwierigkeit unübersehbar, von der Vielfalt individueller Informationslagen und Transaktionen auf die Aggregationsebene des Marktes zu gelangen. Ein notierter Kontraktpreis stellt gewissermaßen die Resultante eines Polzgons unspezifizierter, auf unterschiedlichen Informationslagen basierender Erwartungen dar. Vermutet kann werden, daß seine Informationseffizienz um so höher ist, je zahlreicher und agiler die Marktteilnehmer sind; hier fügt sich die Erfahrung ein, daß verzerrte Kontraktpreise bislang nur auf nach dem Volumen und Partizipation engen Märkten beobachtet werden konnten. Unter Modellierungsaspekten scheint mit der hier vertretenen Sicht des Marktgeschehens noch am ehesten der Versuch vertretbar, zumindest nach Transaktionen und eingenommenen Marktseiten zu unterscheiden,

auch wenn dabei Erwartungen immer noch unzulässig aggregiert werden müssen[28].

Jedoch führt insgesamt wohl kein Weg an der Erkenntnis vorbei, daß das Marktganze eine andere Informationsqualität hat als Ergebnisse von Versuchen, die Vielfalt der Erwartungen der einzelnen Marktteilnehmer zu aggregieren. Das Aggregationsproblem scheint ebenso unlösbar wie in der Gastheorie die Zusammenfassung der einzelnen Molekularbewegungen, die sich jedoch mit den Gasgesetzen erübrigt. Mit dem nicht nachvollziehbaren Qualitätssprung, welcher die Individual- von der Marktebene trennt, läßt sich vermutlich auch ein Teil der Attraktivität rationaler Erwartungen erklären. Die in den Grenzen ökonometrischer Signifikanzkonventionen beeindruckende Informationseffizienz vieler organisierter Märkte suggeriert eine Als-Ob-Vermutung: Es sieht so aus, als ob die Marktteilnemer auf der Grundlage rationaler Erwartungen handelten. Damit erschöpft sich aber schon die Plausibilität des Erklärungsanspruchs dieser Hypothese. Jeder Versuch, sie auf das Verhalten der einzelnen Marktteilnemer zu übertragen, ohne zumindest unterschiedliche Informationslagen sowie davon abgeleitete, internalisierbare Prognosevorteile zuzulassen und damit doch wiederum auf das Aggregationsproblem zu stoßen, muß notwendig an der Frage scheitern, warum es unter diesen Umständen überhaupt zu spekulativen Transaktionen kommt.

Anhand von Terminkontraktmärkten als Beispiel für spekulative Märkte wird gezeigt, daß (1) die Zeitstruktur, die für alle Transaktionen auf solchen Märkten charakteristisch ist, impliziert, daß vor allem zu "falschen Preisen" gehandelt wird, bezogen auf die zugrundeliegenden Erwartungen. (2) Das Zusammenspiel einer Vielzahl von Transaktionsarten ist der Ausdruck der Arbeitsteilung bei den Informationsaktivitäten. (3) Die Informationsqualität der Preise wird durch die Suche nach und die Verwendung von möglichen Informationsvorteilen seitens der Marktteilnehmer bestimmt sowie von der Geschwindigkeit, mit der diese durch Wettbewerb zunichte gemacht werden. (4) Wege, den Schritt von den Entscheidungen der Marktteilnehmer zum Marktganzen und umgekehrt zu modellieren, erwiesen sich bis jetzt als ungangbar ohne die Zuhilfenahme unangemessener Abstraktionen.

## Summary

Heterogenous Expectations, Price Formation, and Informational Efficiency in Speculative Markets

Using futures markets as prototypes of speculative markets, it is argued (1) that the time structure characteristic to all transactions on such markets predominantly implies trading on false prices in terms of underlying expectations. (2) The interplay of the multitude of types of transac-

tions reflects a division of labour in information activities. (3) The informational quality of prices dependes on the traders' search for and use of possible informational advantages and the speed with which these are eroded by competition. (4) To model the step from the decisions of the individual traders to the market level and *vice versa* so far proved to be impossible without resorting to inadequate abstrations.

## Literatur

Danthine, J.-P. (1978), "Information, Futures Prices and Stabilizing Speculation", *Journal of Economic Theory*, 17, 79-98.
Fama, E.F. (1970), "Efficient Capital Markets: A Review of Theory and Empirical Work", *Journal of Finance*, 25, 383-417.
Goss, B.A. and Giles, D.E.A. (1981), *Comparative Modelling of Price Determination and Storage in United States and Australian Commodity Markets: Soybeans and Wool*, Paper presented to the Western Economic Association International Conference, San Francisco, July 2-6.
Grossmann S.J. (1977), "The Existence of Futures Markets, Noisy Rational Expectations and Informationa Externalities", *Review of Economics Studies, 44, 431-449.*
Hawtrey, R.G. (1939/40), "Mr. Kaldor on the Forward Market", *Review of Economic Studies*, 7, 202-205.
Hellwig, M.F. (1980), "On the Aggregation of Information in Competitive Markets", *Journal of Economic Theory*, 22, 477-498.
Hirshleifer, J. (1975), "Speculation and Equilibrium: Information, Risk, and Markets", *Quarterly Journal of Economics*, 84, 519-542.
Johnson, L.L. (1960), "The Theory of Hedging and Speculation in Commodity Futures", *Review of Economic Studies*, 27, 139-151.
Kaldor, N. (1939/40a), "Speculation and Economic Stability", *Review of Economic Studies*, 7, 1-27.
– (1939/40b), "A Note on the Theory of the Forward Market", *Review of Economics Studies*, 7, 196-201.
Kirzner, I.M. (1981), "The 'Austrian' Perspective on the Crisis", S. 111-122, in: D. Bell und I. Kristol (eds.), *The Crisis in Economic Theory*, New York.
Lachmann, L.M. (1943), "The Role of Expectations in Economics", Economica, 10, wiederabgedruckt, S. 65-80, in: L.M. Lachmann (edited with an introduction by W.E. Grinder), (1977), *Capital, Expectations, and the Market Process. Essays on the Theory of the Market Economy*, Menlo Park, California.
Peston, M.H. and Yamey, B.S. (1960), "Inter-Temporal Price Relationships with Forward Markets: A Method of Analysis", *Economia*, 27-28, 355-367..
Rubinstein, M. (1975), "Securities Market Efficiency in an Arrow-Debreu Economy", *American Economic Review*, 65, 812-824.
Streit, M.E. (1980), "Zur Funktionsweise von Terminkontraktmärkten", *Jahrbücher für Nationalökonomie und Statistik*, 195/6, 533-549.
– (1982) *Information Processing in Futures Markets – An essay on the adequacy of abstraction*, European University Institute, Florecne, Working Paper No. 9.
– and Quick, R. (1982), "Informationsaktivitäten und Preisbildungsprozesse auf Terminkontraktmärkten", Jahrbücher für nationalökonomie und Statistik, 197, 309-327.
Tobin, J. (1980), *Asset Accumulation and Economic Activity – Reflections on Contemporary Macroeconomic Theory*, Oxford.

## Anmerkungen

* Für hilfreiche Kritik danke ich Hans Geiger, Alexander Gross und Klaus Kobold (Florenz) sowie Ehrentraud Graw (Mannheim). Alle verbleibenen Unzulänglichkeiten gehen selbstverständlich zu meinen Lasten.

[1] Eine Darstellung der transaktionsbezogenen Funktionsweise dieser Märkte unter Berücksichtigung der institutionellen Bedingungen gibt der Verfasser an anderer Stelle (Streit (1980)).

[2] Kaldor (1939/40b) und Hawtrey (1939/40).

[3] Kaldor (1939/40b), S. 3ff.

[4] Kaldor (1939/40b), S. 196.

[5] Hawtrey (1939/40), S. 205.

[6] Selbst wenn aufgrund dieses Beweises nicht ausgeschlossen werden könnte, daß Spekulanten ihre Marktteilnahme von einer Risikoprämie abhängig machen, dürften sie dies kaum ohne gleichzeitige Bewertung der beobachtbaren Varianz der Prämie tun. Eine konventionell statistisch signifikante Verzerrung, deren Nettoertragswert außerdem größer als der einer risikoarmen Anlage sein müßte, wäre daher immer noch kein hinreichender Beweis.

[7] Eine eingehendere, kritische Analyse dieser Versuche unternimmt der Verfasser an anderer Stelle (Streit (1982)).

[8] Im Hinblick auf Terminkontraktmärkte mag als Beispiel für eine Serie verwandter Modelle das von Grossmann (1977) dienen.

[9] Die gegenteilige Argumentation von Danthine (1978) dürfte nicht schlüssig sein. Er zeigt zwar, daß die Stückgewinne der Risiko übernehmenden Spekulanten durch Informationsaktivitäten in Grenzen verbessert werden können, nicht jedoch, wie der einzelne Spekulant informationsbedingte Prognosevorteile internalisieren kann. Schwarzfahrerverhalten von Spekulanten ist daher nicht ausgeschlossen, was die Aufnahme von Informationsaktivitäten wiederum in Frage stellt.

[10] Dazu gehört auch die Möglichkeit, daß der Terminkontraktpreis weniger stark als der Kassapreis fällt; denn dann läßt sich der Verlust auf dem Kassamarkt wenigsten teilweise durch einen Gewinn auf dem Kontraktmarkt kompensieren.

[11] Die Konstanz von Kontraktpreis bzw. Basis bezieht sich nur auf die Marktsituation zu Beginn und Abschluß des Hedgings, nicht auf die Zwischenzeit.

[12] Die Selektionsmöglichkeiten sind vielfältiger als im folgenden unterstellt. Sie schließen gezielte Abweichungen im Volumen der Terminkontraktposition von dem der Kassaposition ebenso ein wie Abweichungen in den Fristen von Kassa- und Terminkontraktpositionen. Eine theoretische Begründung diskretionären Hedgings im allgemeinen und der zuvor genannten Abweichungen im besonderen wird durch Portfolio Ansatz trotz der unumgänglichen Vorgabe einer Nutzenfunktion erleichtert (vgl. z.B. einen der ersten Versuche von Johnson (1960)).

[13] Aus Gründen der Präsentation wird auch in allen folgenden Ausführungen von der Möglichkeit abstrahiert, daß die geplante Beendigung eines Hedgings nicht mit der Fälligkeit eines Kontraktes übereinstimmt. In diesem Fall würde ein Hedger den Kontraktpreis entsprechend beurteilen müssen, womit sich jedoch die Gründe für vom herrschenden Kontraktpreis abweichende Erwartungen eher noch vermehren dürften.

[14] Eine weitere, empirisch relevante Möglichkeit ist, daß organisatorische Gründe selektivem Hedging entgegenstehen. Die für eine derartige Marktteilnahme erforderliche Dispositionsfreiheit ist z.B. nicht ohne weiteres mit den starren Regeln von Unternehmensbürokratien in Einklang zu bringen.

[15] Diesen elementaren, bei vielen Modellierungsversuchen jedoch vernachlässigten Tatbestand betont besonders Hirshleifer (1975). Jedoch weichen die weiteren Ausführungen von denen Hirshleifers ab, der sog. 'consensus beliefs' als fiktives Bezugssystem (eine

axiomatische Erläuterung des Systems gibt Rubinstein (1975)) benutzt, um heterogene Erwartungen ('divergent beliefs') zu definieren. Das ermöglicht es ihm, eine Variante rationaler Erwartungen beizubehalten.

[16] Was ohne Hedging eine Spekulation auf dem Kassamarkt wäre, wird durch eine entgegengerichtete Spekulation auf dem Terminkontraktmarkt ergänzt.

[17] Damit wird kein Kontinuum hinsichtlich der Verringerung von Ungewißheit im Zeitablauf unterstellt. Die folgenden überlegungen sind z.B. auch auf die korrekte Antizipation transitorischer Störungen der Kontraktpreisentwicklung anwendbar.

[18] Darüber hinaus läßt sich grundsätzlich mit Lachmann (1943), S. 78, vermuten, daß hohe Umsätze bei geringen Preisbewegungen ein besonders deutliches Symptom heterogener Erwartungen sind.

[19] Im Falle der Arbitrage mag das nicht so deutlich sein. Jedoch stellt z.B. intertemporale Arbitrage darauf ab, daß etwa ein herrschendes Contango die Lagerkosten i.w.S. übersteigt. Immer dann wenn die im Zuge der induzierten Arbitrage einsetzende Verringerung des Contangos auch den Kontraktpreis verändert – und das dürfte die Regel sein -, läßt sich argumentieren, daß der Arbitrage zumindest de facto abweichende Preiserwartungen zugrunde lagen.

[20] Selbst in diesem Fall ist nicht sicher, ob ihr Handeln nicht wiederum auf einer abweichenden Erwartung beruht, etwa der, daß sich die Preisbewegung umkehrt und weiteres Halten der Position gewinnschmälernd sein würde.

[21] Einer der ersten, der auf diese Funktion von Scalpern hingewisen hat, dürfte Hawtrey (1939/40), S. 204, gewesen sein.

[22] Auf realen Märkten handelt es sich dabei i.d.R. um Vorgänge im Verlauf eines einzigen Börsentages.

[23] Eine breitere Darstellung des folgenden findet sich bei Streit und Quick (1982).

[24] Die Spezialisierung kann sogar so weit gehen, daß sie sich beim Spreading auf wenige spezifische Termine beschränkt.

[25] Die Informationsanreicherung bewirkt, daß sich die Qualität des Kontraktpreises als Knappheitindikator verbessert. Die der Verbesserung zugrunde liegenden Informationen können jedoch von Marktteilnehmern nicht identifiziert werden. Jedenfalls dürfte die gegenteilige Annahme, die sich in zahlreichen Modellierungsversuchen kombiniert mit rationalen Erwartungen findet (*revealing full information equilibrium*), inadäquat sein.

[26] Dieses Argument betont z.B. Hellwig (1980), S. 492f.

[27] FAMA (1970), S. 387 f.

[28] Vgl. hierzu z.B. den ökonomischen Versuch von Goss und Giles (1981), der theoretisch durch einen früheren Ansatz von Penton und Zamez (1960) angeregt wurde.

# The Knight And His Utility Function*

K. VELUPILLAI
Copenhagen/Florence

> There is no religious denomination in which the misuse of metaphysical expressions has been responsible for so much sin as it has in mathematics.[1]

This essay is an interpretation of a lost manuscript believed to have been written, originally, by the later followers of William of Occam. To put the essay in perspective the following remarks[2] may be useful:

Firstly,

> Occam is best known for a maxim which is not to be found in his works, but has acquired the name of "Occam's razor". This maxim says: "Entities are not to be multiplied without necessity". Although he did not say this, he said something which has much the same effect, namely: "It is vain to do with more what can be done with fewer". That is to say, if everything in some science can be interpreted without assuming this or that hypothetical entity, there is no ground for assuming it".
>
> (Bertrand Russell: "History of Western Philosophy" George Allen and Unwin Ltd., London, 1961, pp 462-463).

Secondly, it is clear that this maxim is applied with admirable consistency by New Classical Economists (cf., for e.g., Thomas Sargent and C.A. Sims: "Business Cycle Modelling Without Pretending to Have Too Much A Priori Economic Theory." In NEW METHODS IN BUSINESS CYCLE RESEARCH: PROCEEDINGS FROM A CONFERENCE, edited by C.A. Sims, Minneapolis: Federal Reserve Bank of Minneapolis, 1977). It is also the one unifying maxim between the mnetarists and the New Classicals – at least methodologically. The Monetarists – and all blind applications of classical statistical methods – are, of course, the official purveyors of the Logic of Refutation (cf. below for a loose definition) whereby assumptions do not even matter; only predictions are relevant.

Finally, a lecture, based on this manuscript was delivered on St. Valentines day (14th February 1985) to the Research students of the department of economics in the European University Institute in Florence. At least one of the foot notes and a part of the last few lines of the text may make some sense if these facts are remembered.

## Introduction

*Go on, believe! It does no harm*

Denis Robertson made Alice famous for (literary) economists. He may have been inspired to do so because of the view he had, through the looking glass world of Alice, of J.H. Clapham's "Empty Economic Boxes" (EJ, 1922). There may have been other reasons – but in the eternal *as if* world of economists this is a sufficient reason. However, even he did not realize that the Knight and Alice had much more to do with economics than Humpty Dumpty, Topsy or the Duchess. In fact it is surprising that so little is known about Lewis Caroll's adaption, from economics of various profound observations. A typical example is the following:

"The name of the song is called 'HADDOCKS EYES'." "Oh, that's the name of the song, is it?" Alice said, trying to look interested.

"No, you don't understand," the Knight said, looking a little vexed. "That's what the name is CALLED. The name really is, 'THE AGED AGED MAN'."

"Then I ought to have said 'That's what the SONG is called'?" Alice corrected herself.

"No, you oughtn't: that's quite another thing! The SONG is called 'WAYS AND MEANS'; but that's only what it's CALLED, you know!"

"Well, what is the song, then?" said Alice, who was by this time completely bewildered.

"I was coming to that," the Knight said. "The song really is 'A-SITTING ON A GATE': and the tune's my own invention."

In fact the original was slightly more complicated. I was fortunate to discover a lost manuscript (unknown author) whilst rebuilding my cellar in Svanshall (Sweden). The clue to the deciphering of this valuable and long-lost manuscript was, in fact, provided by the fact that I had been reading about Alan Turing's work at Bletchley Park and Sir Denis Robertson's collection of essays on 'Utility and All That'. The important hint, for deciphering the hieroglyphics of the manuscript, came from re-interpreting New Classical Economics as a combination of Platonism and Revelation.[3] The Platonic part had been evident ever since I realized that Libertarian philosophy in its New Classical variant depended, ultimately, on the Four Ayyars: Lucas, Sargent, Wallace and Barro. (Ayyar is the Tamil word for the Hindu priests who had the exclusive right to interpret the Hindu scriptures). For those with a Western taste and Greek thoughts it may be useful to point out that Democritus' FOUR kinds of atoms, the ultimate constituents of matter, were Earth, Air, Fire and Water. (Today, of course, they would be called LEPTONS). These were shaped, according to Plato, like FOUR of the regular polyhedrons.

But more importantly, the Platonic part derives from the incredible faith in objectivity and objective values that the world is supposed to be laced with – quite in accordance with Plato's theory of the True Forms. Although lip service is paid to subjective values these latter are, in fact, dispensable artefacts. For, as in the theory of True Forms, the Philos-

opher-Kings – which every economic agent is, at least with respect to the facts of economic life – knowing the existence of the True Forms, are only trying to find them. It is there to be found. The fabric of the world is garnished with objective probability distributions. The economic agents are groping in the cave with the light provided by their subjective probability distributions searching for what they know does exist. But with one fellsweep – the assumption of Rational Expectations – the process of search is banished form the discourse. The Philosopher-Kings are always armed with the TRUE FORMS – the economic agent knows that there is only one True Model of the economy; that is sufficient information for the further assumption that (s)he also has complete knowledge of the one TRUE MODEL.[5]

As for Revelation, and not critical rationalism in the Lakatosian sense (although leading members of the New Classical School of thought claim to follow a Research Programme in the sense of Lakatos), this was evident *not only* from their total commitment to *belief* in a) MODUS PONENS, b) MODUS TOLLENS and c) the law of the Excluded Middle[6], but also from their total lack of intellectual culture. The absence of intellectual culture or, what comes to the same thing, the blind committment to a methodology based on a crude form of positivism – the logic of refutation, which is base on (a)ß(c) above – is in fact crucial in preventing defection from the fold. In fact one of the original Ayyars (Corresponding to yet another of the Platonic regular polyhedrons) left the fold precisely because his intellectual horizons widened[7]. Like all faiths based on Revelation the New Classicals have their Trinity: Adam Smith, Karl Popper and Frank Knight. The New Classical Models are implemented with Rational self-seeking agents in a competitive environment peppered with sources of "Risk, Uncertainty and Profit" where they – the agents conduct content-increasing falsifiable experiments. Thereby hangs the tale of the Butcher, the Baker and the Brewer.[8] That others, like Keynes, had visions about 'Risk, Uncertainty and Profit' is dismissed by an appeal to faith, ignorance and dogma.[9] Platonism and Revelation are cemented by the ubiquitous principle of maximization (or minimization) – i.e., choice theory. Somewhere, somehow, someone must optimize something if whatever is written is to be read, refered and published. A giant, on whose shoulders the Ayyars stand, summarized it all in his Nobel Prize Lecture entitled: "Maximum Principles in Analytical Economics" (Paul Samuelson, Nobel Memorial Lecture, December 11, 1970). To illustrate the twin concepts involved in choice theory and its empirical counterpart, demand analysis, Samuelson invoked the famous analogy of Fermat's least time principle:

> The very name of my subject, economics, suggests economizing or maximizing. . ..(A)t the very foundations of our subject maximization is involved. . . What is it then that the scientists finds useful in being able to relate a positive description of behaviour to the

solutions of a maximizing problem?. . . My positive descriptive relations would be interpreted as the necessary and sufficient conditions of a well defined maximum problem.
(Paul Samuelson, op.cit. pp 130-131).

Paul Samuelson then went on to illustrate the usefulness of this variant of Occam's maxim in Physics by describing the least-time principle of Fermat for a beam of light subject to reflection. (cf. also Paul A. Samuelson: "A Catenary Turnpike Theorem Involving Consumption and the Golden Rule", AER, June, 1965).

But are these analogies useful-indeed are they even up to date from the point of view of developments in theoretical physics? Or are we still apeing Newtonian Mechanics and the concomitant Mechanistic Philosophies? The late Jacksonian Professor of Physics in the University of Cambridge, Otto Frisch, one of the architects of "fission" and indeed the man who coined the phrase 'nuclear fission' was perhaps more to the point – at least so far as the usefulness of analogies are concerned for economists and economics:

For instance, a beam of light passing through refracting media. . . was shown to travel along a path that ***requires the minimum time according to the wave theory of light.*** But the basic law of refraction can be deduced without reference to that parsimonious principle and, moreover, allows light to travel equally on a path that requires ***not the least but the most time*** (at least compared to all neighbouring pathways). ***Today such minimum (or maximum)*** principles are "regarded merely as pretty (and sometimes useful)" consequences of more basic laws. . .
(Otto Frisch: "Why" in *The Encyclopaedia of Ignorance*, Pergamon Press, 1977; Italics Added).

For New- and Neo- and every other type of macro- and microeconomics, except in the 'underworlds' of deviants,[10] parsimony and optimization are 'not merely pretty' or 'sometimes useful'; they are the bread and butter of the subject. Indeed the death of Macroeconomics was recently announced (cf. Financial Times, 19/10/84) and the disease which precipitated the untimely death was chronic ad-hockery (or lack of sufficient Optimization Cells). It was not for lack of understanding of this sickness by eminent macroeconomists; indeed none other than a distinguished recent Nobel Laureate diagnosed it better than most antagonists:

One view, prevalent among mathematical theorists of general equilibrium, is that traditional macroeconomic theory suffers from its lack of firm microeconomic foundations. *The behavioural relations of macromodels, it is said, are not rigorously derived from optimization by individual agents and from the clearing of markets in which optimizing agents participate. In short macro models do not look like general equilibrium models.*
(J. Tobin: Asset Accumulation And Economic Activity, Basil Blackwell, Oxford, 1980, p.x; Italics Added).

But if we take Otto Frisch's "Why's" somewhat seriously should we not be more sceptical about the 'restrictions' derived from maximization

postulates and the parsimony forced upon all and sundry by black-box methods? Clapham's 'empty economic boxes' have been completely and comprehensively replaced by sophisticated 'black boxes' driven by auto-regressive processes with, as a final ironic twist in this game, white noise playing an all too important role.[11]

Should we not be a little more modest about the basic message of the 'Correspondence Principle' a là Samuelson? (cf., my: 'A Note on the Origin of the Correspondence Principle' in "The Swedish Journal of Economics", 1973).

These doubts may pass off – and have indeed passed off despite the pleas of eminent heretics – as water off a duck's back. Nothing in the manuscript I found indicates the slightest doubt in the correctness of carrying the master's (Occam's) maxim to its ultimate and ludicrous ends.

The next few paragraphs transcribe the part of the manuscript that I have, so far, been able to decipher. I expect to be able to give a fuller version in due course.

## Text

"Don't 'for heaven's sake', be afraid of talking nonsense! But you must pay attention to your nonsense."

"The name of the GAME IS CALLED 'THE NEW MACROECONOMICS'."

"Oh, that's the name of the GAME, is it?" Alice said, trying to look interested.

"No, you don't understand," the Knight said, looking a little vexed. "That's what the name is CALLED. The name really is, 'NEW CLASSICAL ECONOMICS'."

"Then I ought to have said 'That's what the GAME is called'?" Alice corrected herself.

"No, you oughtn't: that's quite another thing! The GAME is called 'EQUILIBRIUM BUSINESS CYCLES'; but that's only what it's CALLED, you know!" "Well, what is the GAME, then?" said Alice, who was by this time completely bewildered.

"I was coming to that," the Knight said. "The GAME really is 'RATIONAL EXPECTATIONS': and the rules are my own invention."

At this point, hearing that the Knight had said: ". . .the rules are my own invention. . .", his Utility Function started stirring from its deep slumber. Gradually the stirring became a rumbling and then a very clear protest. "The rules", his Utility Function told him, "have always been mine.[12] I was born with them."

"I have allowed you to BEHAVE at different levels of consistency. In return you have not driven me too hard. You have respected my CIRCADIAN RHYTHMS – my biological clock, "the dynamics of my physiological rhythms". You have been very wise to respect such dynamics although I suspect that you have simply ignored them. Your 'theoretical technology' is not quite up-to-date to study my CIRCADIAN RHYTHMS, (cf. Arthur T. Winfree: The Geometry of Biological Time, Springer-Verlag, Berlin, 1980; esp. chapter 19).

This came as a shock for the Knight. He had, all along been under the impression that it was he who was being RATIONAL. He suddenly realized not only that his RATIONALITY was indissolubly linked to the Utility Function – but also and more disturbingly, that he THE Knight, was incapable of analyzing DYNAMICS and RATIONALITY within the same framework.[13] This the Knight was supposed to gather and deliver to the Utility Function which then went about optimizing. It is in gathering – acquiring – the information that the Knight was allowed to invent rules – or so he thought.

"I beg your pardon," the perturbed Knight said, now, to his Utility Function:

"I meant that the GAME of RATIONAL EXPECTATIONS is played by myself according to RULES I have invented myself. The rules are about little bits[14] called information. I arrange these bits of information according to rules I have invented – and I collect such bits; sometimes I buy them – other times I exchange my bits for bits owned by others. There is a market for bits of information."

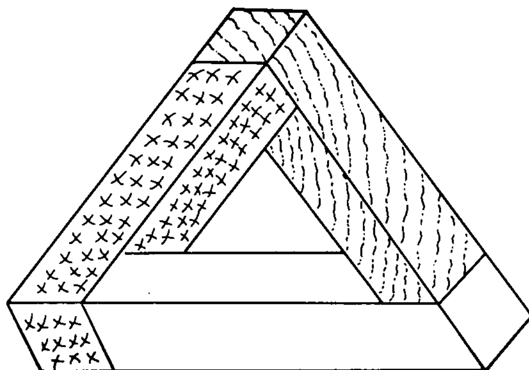
At this point the Knight became a little confused. He suddenly realized that his "rules" were actually, once again, those of the Utility Function. He remembered, too late, the remarkable words of the most venerable of Ayyars:

"The assumption of rational expectations is appealing (in the context or yielding important insights into the workings of market economies) not only because *by applying the principle of opimization to the acquisition of information* it maintains an internally consistent framework, but also because it limits to some extent the range of possible outcomes."

(R.E. Lucas, Jr., Review of 'Money and Value: A Reconsideration of Classical and Neo Classical Monetary Theories" by Jean-Michel Grandmont, in: The Journal of Economic Literature, vol. XXII, n. 4, p. 1653; Italics Added).

Thus, in buying and selling such bits of information – i.e. in the acquisition of information – he was being Rational – according to the rules imposed by his Utility Function. Was he anything? Did this mean that the 'Turing Test'[15] was about Utility Functions and their Identification! Was the business of being Rational leading to familiar problems of infinite regress[16] – especially if he – the Knight – had to take into consideration the activities of Bishops, Pawns. . .? Expectations of Expectations; Rationality of Rational Agents and Rationality of Rationality! If he was going to be consistent did he have to be 'incomplete' or was it the other way about – or both! Having just recalled "Turing's Test" he was now tormented by Godel's theorem.[17] These fleeting, disturbing and intellectually honest worries were dispatched summarily to the innermost recesses of 'mind', equilibrium and poise were restored and the Knight bounced back with a spectacular realization. He suddenly realized that, in fact, his RULES for the GAME were something different altogether. The

GAME really was RATIONAL EXPECTATIONS. But the RULES were about BELIEFS which, in turn, depended on PERCEPTION. He hastily corrected himself. "The GAME really is RATIONAL EXPECTATIONS. The rules are for the bits of information which I value according to my BELIEFS which, in turn, are based on my PERCEPTIONS.[18] He was now feeling exhilerated at being one-up on his Utility Function. "Indeed," said his Utility Function, "could you tell me what you BELIEVE when you PERCEIVE the Reutesvard (-Penrose) triangle:[19]



The Knight was particularly pleased when his Utility Function asked him to interpret the Reutesvard (-Penrose) 'impossible triangle'. In fact the Knight had just been reading the late David Marr's new book on 'Vision' precisely to help in understanding the paradoxes posed by Impossible Diagrams, Fractals and other Chaotic objects. He was, therefore, able to quote almost verbatim from Marr:

"Examples like the Penrose triangle, many of Escher's figures, . . ., probably depend on a mixture of effects (or internal computations), some local in the 2 1/2 D sketch, and other effects due to a failure to construct an overall, consistent three-dimensional interpretation from a set of local views."

(David Marr: 'Vision: A Computational Investigation into the Human Representation and Processing of Visual Information'., W.H. Freeman and Co., San Francisco, 1982; p. 294).

The Knight had been completely convinced by this new view of vision. His Utility Function was not totally ignorant about matters relating to recent developments in the theory of vision. After all, most of its hard work had been to disentangle the illusions of its Knight and these were not only the conventional hallucinations. In fact most of the illusions had come about after the Knight had got one of his Sargents to do a study on

various forms of 'Observational Equivalences' (cf. op.cit. Sargent-Sims) and, even more importantly, after Phelps had suggested the famous 'Island Paradigm'. After the latter suggestion most New Classical Knights moved lock, stock, Bishop, Rook and Pawn to various islands. A fixed total supply of the Financial Times, the Wall Street Journal, Money and Manna was then randomly distributed to the various islands. Since all modern forms of communications were unavailable, prohibited or ignored, the acquisition of information of the random distribution to other islands was inferred by ingenious home-made devices. Smoke-signals, conch shells, palm leaves, witch doctors and much else was harnessed for the purpose of accurately distinguishing local information from the global facts. All this had necessitated, on the part of the Utility Function, complex computations and detailed analysis of various forms of continuity[20] manifested in impossible triangles, fractals and strange attractors. The common factor seemed to be that an intelligent interpretation of 'observed' or 'observable' phenomena required an abandonment of the exclusive dependence on Integer Dimensions. The 2 1/2 D of David Marr, the Fractal Geometry of Benoit Mandelbrot and the Hausdorff Dimensions of Strange Attractors had, at least, challenged the tyranny of integers – and, thus, the 'naturalness' of natural numbers. This, perhaps the first real challenge to the Kronecker-Brouwer programme of Intuitionism, if successful, would restore the Mathematical Method Where it 'belongs': in Cantor's paradise of continua, infinite cardinals, etc.[21]

Both the Knight and his Utility Function were, thus, pleased with Marr's tentative explanation for this restored some respectability also to the behavioural assumptions about asymmetric information due to confusions about local and global factors (cf. discussion above) – much (mis-) used in New Classical formulations. The Utility Function had not read Marr's book but had indeed read the Scientific American article by one of Marr's close collaborators, Tomaso Poggio (Scientific American, 1984) and the review article on Marr's book by Israel Rosenfeld, in the London Review of Books (October 11, 1984).[22]

The Utility Function very gently asked his Knight whether it was not true that the Marr-Poggio interpretation was not a computational theory of vision. The Knight answered in the affirmative – but, now, with the same sense of unease as he had felt a little earlier when speaking about bits of information. His unease was well-founded. The Utility Function then, even more gently, enquired whether this did not mean that the Knight was, in fact, totally irrelevant: A Universal Turing Machine with Utility Functions as inputs was sufficient!!

The Knight was, by now, completely bewildered. He had one last line of defence for his integrity and existence. He decided to take the initiative and ask questions himself. His first was about the setting in which behaviour was supposed to take place. After all, he was (Frank) Knight.

Competition was the backdrop against which (rational) behaviour was supposed to take place. Without some form of competition what was the point of the Utility Function maximizing anything. Indeed, there was no role for the Utility Function if not for the fact that there was like Aether, competition – in some form – permeating the medium in which 'it' was to operate.

"It is not true", the Knight asked his Utility Function, "that for you to exist it is essential that some form of competition had to be postulated. Indeed, for simplicity, is it not true that pure uncontaminated competition is the best medium for you, the Utility Function, to operate?"

The Utility Function's reply was most learned. "Yes", answered the Utility Function, "it is true that competitive behaviour is the rationale for my existence. But by that token you become even more insignificant – indeed you do not exist. For, you see, the truly rigorous way of justifying the competitive principle is by an appeal to a continuum of traders and markets as Aumann and others have shown. In such cases you as an individual become meaningless. On the other hand you can be considered "an infinitesmal." That, however, for rigorous justification requires Non Standard Analysis.[23] There really is no hope for you. If you are 'large' competition is useless and I cannot be of much use to you. If you are 'small' you must be 'really small' and competition means you must, in fact, be an 'infinitesmal' and then I can exist but you make no sense. It is sufficient with a UTM.[24] The real problem is to locate Maxwell's Demon (the Auctioneer) to crank the UTM."

The Knight had, by this time, given up all hope. He remembered the words of a great sage Rabbi Hillel:[25]

> "If I am not for myself, then who is for me?
> And if I am not for others, then who am I?
> And if not now, when?"

Indeed Rabbi Hillel knew much more about the problem of coordinating *selfish* and *altruistic genes*, statics and dynamics and micro- and macro-economics. He, in fact, had done his homework – which means: read Hicks:

> ". . . (in) discussing the microeconomic foundations of macroeconomics. . . we had come to realize that there were several kinds of macroeconomics, each probably requiring its own foundations."
>
> (Sir John Hicks: "Final discussion" in 'the Microeconomic Foundations of Macroeconomics' G.C. Harcourt., ed.) Macmillan, London, 1977; p. 373.

Eventually, the KNIGHT, his UTILITY FUNCTION, his BELIEFS, his PERCEPTION and ALICE were sent to various seats of learning. As a matter of fact the KNIGHT was sent to Chicago, his UTILITY FUNC-

TION to California, his BELIEFS to Pittsburg (Carnegie-Mellon), and his PERCEPTION to Boston (MIT) all of them to do PhD's. ALICE WENT TO FLORENCE – the EUI or the LOOKING GLASS UNIVERSITY. The darkness shed by Chicago was the KNIGHT's suit.

## Postscript

*If people did not sometimes do silly things, nothing intelligent would ever get done.*

The most venerable of Ayyars, when reflecting upon methodological and other weighty and mighty issues of philosophy, gave us a hint on *how* to read the writings of Occam's disciples:

"Economist who find Keynes's style congenial will continue to use his writings as Denis Robertson did Lewis Carroll's, but surely there is more to the cumulative nature of economics than this!"

(R.E. Lucas, Jr.: "Methods and Problems in Business-Cycle Theory" in 'Studies in Business-Cycle Theory' by Robert E. Lucas, Jr.; Basil Blackwell, Oxford, 1981; p. 276).

The attempt to decipher the 'lost manuscript' was facilitated by a literal interpretation of the above hint. The real message may well be that *all* writings by economists should be used the way Denis Robertson did Lewis Carroll's!

The central theme of my own interpretation of the 'lost manuscript', on the other hand, is (with due apologies to W.H. Auden and Christopher Isherwood):

"Let us honour, if we can, the Reasonable Man; Although we value none but the Rational Man."

- in Economics, of course.

## Notes

* Absolute fiction. Please do not quote with or without permission. These notes were conceived while waiting for a vaporetto to take me from Giudecca to San Marco in Venice on a lovely morning in early February. The first version was written, a few days later, on a terrace set amongst the gentle hills of Fiesole whilst negotiating an exquisite bottle of the 1979 Solaia. The final version was completed whilst on a boat at Elsinore, with a spectacular sunset setting in relief Hamlet's Castle, on my way to Copenhagen. The nature and contents of the paper and the underlying "spirit" reflect the strange road traversed from time it was conceived to its completion. It is humbly dedicated to Andrea, Giorgio and Stefano.

[1] The quotes in the beginning of each section are all from Wittgenstein's "Culture and Value". (L. Wittgenstein: Culture and Value, translated by Peter Winch, Basil Blackwell, Oxford, 1980).

[2] Some of the 'labels' will become clearer in the discussion below.

[3] I would not have been able to understand the significance of 'Utility and All That' if I had

not remembered the opening sentences of Professor Amartya Sen's inaugural lecture delivered at the London School of Economics:

"*Thirty five years have passed since Paul Samuelson published in the house Journal of the London School of Economics his pioneering contribution to the theory of 'revealed preference'. The term was perhaps not altogether a fortunate one. Revelation conveys something rather dramatic, and the biblical association indued the late Sir Denis Robertson to wonder 'to some latter day saint in some new Patmos off the coast of Massachusetts, the final solution to all these mysteries had been revealed in a new apocalypse' (Utility and All That, p. 19). While the appropriateness of the terminology may be debatable the approach of revealed preference has gradually taken hold of choice theory.*"

(Economica, August 1973, p. 241)

Sir Denis slipped on a very minor point. It was a 'new Patmos' off the coast of Lake Michigan (not 'off the coast of Massachusetts') where *on Valentine's Day* all was revealed to Knight and Bliss ruled.[4]

[4] F.A. Valentine and G.A. Bliss were pioneers in extending classical calculus of variations to include inequality constraints: i.e., forerunners to today's 'Optimal Control' or one of the pillars propping up the Lucasian structure called 'Theoretical Technology". Frank Knight is the Patron Saint of New Classical Macroeconomics. They were all from Chicago.

[5] ". . . the rational expectations program of policy analysis logically requires the authority of a single model Hence thorough – going implementation of the rational expectations method in policy making would entail the official promotion, or 'establishment', of *ONE MODEL* over the others."

(Individual Forecasting and Aggregate Outcomes: 'Rational Expectations' Examined by R. Frydman and E.S. Phelps, CUP, 1983; in "Introduction" by R. Drydman E.S. Phelps, p. 27 – Italics Added).

The Platonic TRUE FORM, the 'true model' is simply there to be found. There is, in the new libertarianism, no scope for INVENTING new models. We can, at most, discover. Brouwer's fixed-point theorem (or generalization) is often invoked in using the new "Theoretical technologies' (Lucas' colourful word: cf. R.E. Lucas, jr: STUDIES IN BUSINESS CYCLE THEORY, Basil Blackwell, Oxford, 1981; p. 9) True to the 'revealed' nature of their knowledge, the New Classicals – or, for that matter, most of the Neo-Classicals (with honourable exceptions, e.g., Herbert Scarf) – are unaware that Brouwer was an Intuitionist actively promoting constructive inventions and denouncing discoveries including his own fixed-point theorem.

[6] Again, Brouwer led the Intuitionists in rejecting this assumption in all reasoning involving the uncritical use of the infinite.

[7] Leonard Rapping.

[8] cf. Adam Smith: "The Wealth of Nations", Everyman's Edition, p. 13. For those not quite up-to-date on the Hindu trinity may I enlighten them on the fact that Shiva is the Butcher (the Destroyer), Brahma the Baker (the Creatoer) and Vishnu the Brewer (the Preserver). It must also be mentioned, however, that there are several faiths claiming Smith as one of their Trinity.

[9] Arjo Klamer *Other people say that there are periods with fundamental uncertainty. They refer to the Keynesian concept of uncertanity. As a matter of fact, Rapping tells this story, too.* Robert E. Lucas jr. "*It's Knight not Keynes*" *(as reported in: "The New Classical Macroeconomics; Conversations with New Classical Economists and Their Opponents."* ed. by A. Klamer, Harvester Press, 1984, p.44).

[10] cf. p. 32, of '*The General Theory of Employment, Interest and Money*' by J.M. Keynes (Macmillan, 1936) for a short-list of 'underworld deviants'.

[11] Is it just my fanciful imagination or have the relevant authors consciously chosen words and processes: The EMPTY set is a subset of ALL sets; BLACK (colour) is the equivalent

in that it absorbs everything; and WHITE (colour) can of course be decomposed into everything (and anything)!!

[12] Recall the Utility Function (of a greedy agent) is only a proxy for the more basic Preference Relation. That, in turn, is only a reflection of an even more fundamental entity: The Selfish Gene. Indeed:
"... the fundamental unit of selection, and therefore of self-interest, is not the species, nor the group, nor even, strictly, the individual. It is the gene, the unit of heredity"
*(R. Dawkins: The Selfish Gene, Oxford University Press, 1976; p. 12, Italics Added).*
Of course this does not mean that there are no "Altruistic Genes". It is just that in *orthodox theory – as Professor Hahn likes to remind us every third hour* – we only model the Selfish Gene. In a celebrated article in NEW BIOLOGY (vol 18, pp 34-51) called 'Population Genetics' the famous and eccentric J.B.S. Haldane remarked that he was 'prepared to lay down his life to save two brothers or eight cousins'; the morale of the act being to induce an increase in the frequency of *Altruistic Genes* in his descendents. Such exotic thoughts belong to the 'underworld of deviants' – not to the minds of economic theorists who are wholly conditioned by *the Selfish Gene*.

[13] "I was lucky that I did not too-consistently stick to my program of deriving operationally meaningful theories, for otherwise topics like welfare economics and *economic dynamics might have been excluded*. Though the *accelerator-multiplier interactions cannot usefully be analyzed in terms of any maximum problem*, it would be a shame to leave such models out of an economic discussion."
(Paul Samuelson, Foreword to the Chinese Translation of *Foundations of Economic Analysis*; Italics Added).
and again:
"My point in bringing up the accelerator-multiplier... is that it provides *a typical example of a dynamic system that can in no useful sense be related to a maximum problem.*"
(Paul Samuelson, (loc.cit., p. 141; Italics Added)

[14] An abbreviation of binary digit.

[15] Alan M. Turing: 'Computing Machinery and Intelligence' in Mind, vol. LIX, No. 236, 1950.

[16] "The shadow of the future and of the future's future also affects current outcome. ... Every generation must accumulate ... assets to provide for old age. ... ..their purchasing power in retirement depends on the prices the next generation will pay for them. Those prices in turn will depend on, among other things, the prices the young expect their young will later pay for them. And so on, AD INFINITUM. For certain stores of value. . .there is no intrinsic value and *only on infinite regress of* expectation."
(Tobin, op.cit, pp 26-28; Italics Added)

[17] cf. for e.g.,: Ernest Nagel and James R. Newman: "Godel's Proof" Routledge and Kegan Paul, London, 1959.

[18] As two leading proponents of the New Macroeconomics in, appropriately, the Island Kingdom (recall the importance of the 'Island Paradigm' for the third of the pillars that sustain the Lucasians: asymmetric information or local vs. global confusion; the other two pillars being rational expectations and continuous market clearing) put it:
". . .on this view (hypothesis of rational expectations) behaviour reacts to expected future behaviour which in turn depends on current behaviour; the capacity exists therefore for *changes in the environment to affect current behaviour sharply as individuals react to their perceptions* of the changed environment and its implications for the future. Expectations are therefore completely integrated into behaviour."
*(P. Minford an M.J. Peel: "Rational Expectations and the New Macroeconomics". Martin Robertson, Oxford 1983, pp 4-5, Italics Added).*

[19] This 'impossible triangle' known in the literature as the Penrose Triangle was, in fact, first

conceived by Oscar Reutesvard in the early 30s.

[20] Neither the Knight nor his Utility Function could ever really forget the epitaph (the only possible word after what Sraffa did to him) in Marshall's 'Principles': NATURA NON FACIT SALTUM. Continuity is almost as ubiquitous as optimization (*Rationality). The Rational Agent, the Greedy Individual, the Optimizing (Economic) Man and the Continuous Homo Oeconomicus are all, almost, synonymous words. But in his rush to apply Newton's Infinitesimals and Darwin's 'Survival of the Fittest' Marshall forgot to mention the warning that Darwin's great friend gave. In fact on the very day that Darwin's 'The Origin of the Specis' was published Thomas Huxley wrote him a letter warning Darwin that he was taking a risk in becoming inextricably married to the idea of slow and constant evolution as summed up in the latin tag NATURA NON FACIT SALTUM (SIC!!): "You have loaded yourself with an unnecessary difficulty" wrote Huxley. Darwin was as adamant as Marshall and his disciples.
(cf: F. Hitching: 'The Neck of the Giraffe or where Darwin went wrong, Pan Brooks, London, 1982, p. 17).

[21] Kronecker's famous statement is: "God created the natural numbers. All else is manmade." Intuitionism and the rejection of the law of the excluded middle in Mathematical Proofs relies heavily on the 'naturalness' of natural numbers.

[22] The Utility Function had been directed to Israel Rosenfeld's review article by an Abbruzzese who, although living in Florence, had been unable to persuade *his* Utility Function to leave the Abbruzzo. In fact there is no known instance of an Abbruzze's Utility Function ever leaving that region. The only other people with such strange behaviour are the Tamils of Southern India and Northern and Eastern Sri Lanka.

[23] Although there have been sporadic attempts to apply Non Standard Analysis in Economic Theorizing it has not tickled down to text book levels for ordinary mortals to make any sense of it. At the moment it is not clear whether the Knight is standard or not. The ideas go back to the obscure works of the Italian Mathematician Veronese ('Fondamenti di Geometria', Padua, 1891). The modern revival – even in economics – owes much to the works of the late Abraham Robinson.

[24] UTM: Universal Turing Machine.

[25] Quoted by K.J. Arrow in: 'The Limits of Organization', W.W. Norton & Co., New York, 1974; p. 15.

# IV

# History

# Scholars' Wives, Textile Workers and Female Scholars' Work: Historical Perspectives on Working Women's Lives

GISELA BOCK
Berlin/Florence

## I.

"Woman's place is in the history books," proclaimed Anne Firor Scott in 1979.[1] However, until recently this was, in the bulk of historiography, true only for the final ("last but not least") section of scholars' acknowledgments: "But without Margaret, who suffered in ways well known to scholars' wives, the proofreading would have been interminable and the publication considerably delayed", or: "I want to thank my wife for valuable counsel from first to last; without her self-sacrifice, patience and ever present sense of humour this book could not have been written."[2]

Obviously, though surreptitiously, such gratitude refers to an important aspect of working women's lives: housework, a kind of work on which not only most scholars' happiness, but also their achievements, largely depend; which is often conceived not as work but as love; which is unpaid but secures – though in an unstable way – the support of supporting women. Recent efforts to rewrite history by including not only men's but also women's experience in and of history have underlined the need for a redefinition, from the perspective of women, of many current historical and social categories. Among them was the concept of "women's work" and "working woman" which should include the historical significance and change not only of paid, but also of unpaid, work.[3] In fact, it was estimated that presently 90% of the unpaid work of the world is done by women and that, correspondingly, women's share of the income derived from work is by far inferior than men's.[4] The history of working women's lives is to a large degree the history of female housework.

Despite the current and common invisibility of such women's work, various scholars, mostly economists and sociologists between the 1880s and the 1930s, have noted it. They attributed to it a high individual, social and economic value based not only on its achievements, but precisely on its lack of remuneration. One example is Lorenz von Stein who, in 1886, dedicated his book *Die Frau auf dem Gebiete der Nationalökonomie* to his

"Verehrte Freundin!" His investigation intended to counteract "eine der merkwürdigsten Erscheinungen unseres Jahrhunderts, dies Loslösen der Frau vom Mann, dies Gefühl der Selbständigkeit". To the women who claimed "to be valued according to their value" he demonstrated the value of women's housework ("Arbeit der Liebe") to their husbands, starting from his own experience: "Ich kann freudig in den Tod gehen, wo es das Höchste gilt; aber – und das Triviale mit seiner kalten Stirn wird furchtbar ernst, wenn ich ihm in das herzlose Auge zu scheuen den Muth habe – ich kann diese Begeisterung und diesen Muth nicht fünfundzwanzig Jahre in meiner Arbeit aufrecht halten, wenn ich nirgends eine freundliche Stelle finde, auf welcher mit der Verzehrung zugleich ein wohlthuender Genuß entgegenkommt, wenn Haus und Bett und Tisch und Kleidung, ungemüthlich und unsauber, mir täglich ihre erkältenden Tropfen in den glühenden Becher meiner begeisterten Arbeit gießen. . ." He underlined that "der Mensch (meaning, as is often the case in the German language, 'der Mann'), der in ewig neuer Thätigkeit sich bewegt, kann nicht ohne Ordnung sein. . . Wer aber soll ihm. . . in der Herstellung der Ordnung für Zeit und Raum seines Hauses helfen?" Of course not "der Mensch"[5]: "An der Schwelle dieses Hauses aber steht die Frau. Ich weiß wohl, was ich dort von ihr erwarte; ich weiß, daß ihre weiche Hand mir die Stirn glättet. . . und daß die Arbeit an mich kein Recht mehr hat. . . Und diese Arbeit der Frau ist es, die in ihren tausend kleinen Mühen und Aufgaben doch wieder eins ist, unendlich wie das Leben selbst, aber zuletzt der Werth aller Werthe, die ich gewonnen haben mag." As an economist, he stressed not just the emotional but also the financial value of female work to men: "Und wenn ich nun vom Gefühle zum Verstande übergehe, so wird aus dem was freundlich ist, etwas, was mir mit jedem Jahre mehr auch seinen wirthschaftlichen Werth enthüllt. . . Nehmen Sie einen Augenblick den Stift zur Hand – . . . Und daß das in zehn Jahren einige hundert Millionen gibt, um die wir reicher sind, wenn die Frau des Hauses in wirthschaftlichem Sinne Hausfrau ist?. . . Sie lächeln? Ja, es ist auch komisch, von solchen Dingen überhaupt und noch dazu wissenschaftlich reden zu wollen. Aber doch kann man ja einmal über die Sache nachdenken. . ." The results of this thinking were indeed all but "comical"; he concluded that the essential task of housework consisted in administrating that part of the husband's income – and, according to social class, the lack of it – which was to be used for the household: "Kochen ist in erster Linie rechnen, in zweiter wieder rechnen, und rechnen in dritter Linie. Kochen kann jeder, der es bezahlen kann; daß ich (i.e. the woman) das könne, ist die Hauptsache." Therefore he sustained that "das erste und absolute Prinzip aller Arbeit der Frau das Festhalten an der Summe in der Hauswirthschaft ist, welche der Mann der Frau geben kann. Eine Frau, die an diesem Prinzipe noch zweifelt oder in – verzeihen Sie mir das harte Wort – verbrecherischem Leichtsinn es verletzt, verdient nicht den

hoch ehrenwerthen Namen einer Frau." Precisely for all its value, women's work is "nicht zählbar und meßbar, und doch erreichbar, nicht bezahlbar und käuflich, und doch so unschätzbar", and the author questioned seriously the opinion that "es der Liebe und der Achtung vor unseren Frauen Eintrag thut, wenn man ihnen beweist, daß sie uns nicht bloß unendlich theuer, sondern daß sie uns außerdem auch noch mindestens tausend Millionen, zu sechs Prozent berechnet werth sind."

In his *Grundriß der allgemeinen Volkswirtschaftslehre* Gustav Schmoller continued to think about the matter and expressed his gratitude to "Meiner teuren Frau Lucie, dem Stolze und dem Glücke meines Lebens, der treuen Gefährtin meiner Arbeiten". Later on in the book he specified: "Die Gattin, die dem Manne das Mahl bereitet, ihm abends die Stirn glättet, die Kinder vorführt, wird dienend zur Glück spendenden Herrscherin ihres Hauses. . . (Sie) waltet in Küche, Keller und Kammer, sie reinigt und flickt, stellt überall im Hause wieder Ordnung her, führt den kleinen Kampf gegen Staub und Verderbnis und erhält so allen Besitz, alle Geräte, alle Mobilien sehr viel längere Zeit; sie kann mit demselben Einkommen das Doppelte schaffen, wenn sie ihr Budget richtig einzuteilen, wenn sie mit Waren- und Menschenkenntnis einzukaufen versteht, wenn sie die nötigen kleinen chemischen, technischen und Küchenkenntnisse hat. . . Was macht die Arbeit billig und gut? Daß sie mit Liebe für Mann und Kind, für das eigenste Interesse erfolgt, daß sie nicht bezahlt und gebucht wird, daß dabei nicht gerechnet wird." The essence of the unpaid work done by women of all classes has been seen in the wise administration and the stretching of the husband's income, conceived as – particularly in relation to poor women – "die kleine Sozialpolitik" compared to the "große Sozialpolitik" of wage rises and social security for male workers.[6] Some generations after Lorenz von Stein, there was less "Lächeln" about scholarly thinking on the subject of housework, just as the "Stirn glätten" gave way to more prosaic descriptions: "Die Frau findet in der unmittelbaren Führung der Haushaltungsgeschäfte die in der Regel ihr am meisten zusagende und zugleich die wirtschaftlich fruchtbarste und nützlichste Wirksamkeit."[7]

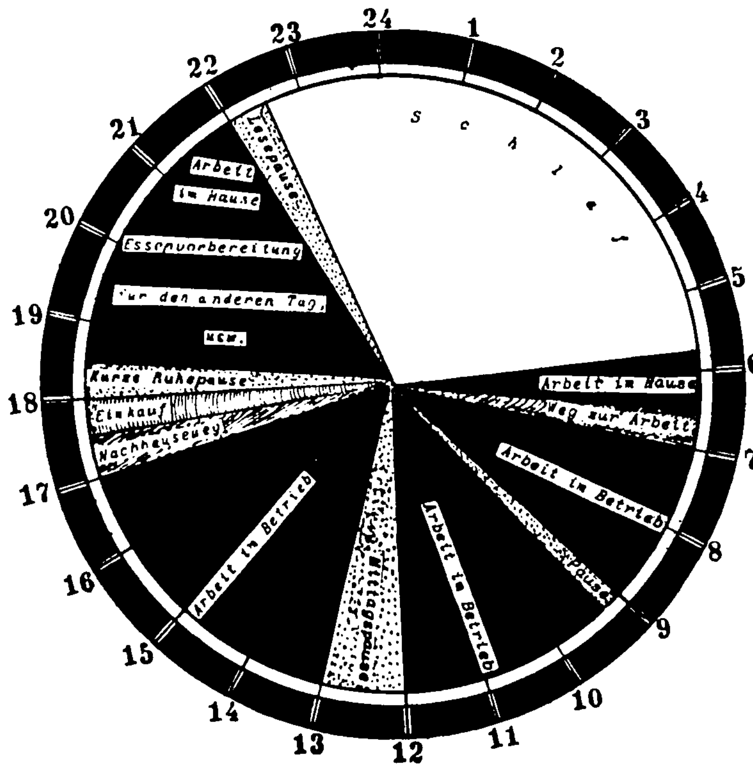
As female housework changes over time, so does scholars' awareness and acknowledgment of it. Two generations later, at a time of renewed doubts about women's proper work and increased access of women to scholarly work, husbands might choose to praise different kinds of female support, such as the author of a book on slaves: "My wife, to whom this book is dedicated, did not type the manuscript, do my research, darn my socks, or do those other wonderful things one reads about in acknowledgments to someone 'without whom this book could not have been written'. Nor did she work so hard on this book that she deserves to be listed as co-author; if she had, she would be listed as co-author. She did, however, take time from writing her doctoral dissertation to criticize each

draft, review painstakingly the materials, help me rewrite awkward sections and rethink akward formulations, and offer countless suggestions, corrections, and revisions. And while under the pressure that anyone who has written a dissertation will readily appreciate, she made an immeasurable if intangible contribution to the writing of this book by living it with me."[8]

## II.

In the last decade, mostly female scholars have dealt with historical continuity and change in the forms, contents and perceptions of housework, with differences in housework according to class or, rather, to the husband's income, with the relations between paid and unpaid work of women, with differences in cultural context and cultural meaning of women's work inside and outside the home.[9] It has been shown that behind the image of the middle-class "lady of leisure" of the turn of the century there was a daily life of work which, moreover, included the effort to keep such work invisible: German "middle-class wives were to appear idle, and by their apparent non-involvement with housework they were to demonstrate and symbolize the social status of their husband."[10] Other authors have argued that for a variety of reasons – such as, e.g., rising standards of cleanliness, of childraising and of support to husbands, or a process that has been called "emotionalization" of housework – the mechanization of the household since the 1920s has hardly diminished women's working hours in the home.[11] The investigation of housework has led to look at the family not just as a unit of mutual interest and complementary dependence but as a workplace for one sex and a place of rest for the other; as a locus not (only) of harmony but also of (potential) conflict[12]; not just as a division of labor between the sexes but also a gender-determined division of independence and leisure, wealth and welfare; decision-making and power[13]; it has questioned the familiar dichotomies and contrapositions of "work and family" (since both are workplaces for women), "production and reproduction", "public and private".

Women's working lives have rather been a daily continuum and interaction of work inside and outside the home, particularly so in the case of working women's lives in the more restricted sense of the term, i.e. employed women or – even more restricted – employed working-class women. Many of them worked outside the home out of necessity, often for the mere survival of their families, as for instance most of the German textile workers who in 1928 responded to a women union's appeal to describe *Mein Arbeitstag – Mein Wochenende*. For them, the workday started long before the factory whistle blew and was not finished after the shift nor at the weekend: "Meine tägliche Arbeitszeit in Haushalt und Fabrik beträgt 16 bis 18 Stunden. Unsere Familie zählt fünf erwachsene




*Der Verlauf des Arbeitstages einer Textilarbeiterin in bildlicher Darstellung*

Köpfe im Alter von 15, 17, 19, 42 und 44 Jahren. Frühmorgens 4.30 ist die Nacht vorbei; dann ist es höchste Zeit, mich anzukleiden, zu waschen und das Frühstück den anderen zu besorgen und alles zu wecken, denn 5.15 heißt es nach dem Betrieb zu gehen, da ich 45 Minuten bestimmt laufen muß. Um 6 Uhr beginnt die Arbeitszeit im Betrieb. Ich arbeite in einer Tuchfabrik und zwar in der Zwirnerei. Zuspätkommen darf im Betrieb nicht vorkommen. Wir arbeiten in Schichten von 6 bis 14 Uhr und von 14 bis 22 Uhr. Arbeite ich in der Vormittagsschicht, so geht von 14 Uhr die häusliche Arbeit an. Ich muß dann auf dem Wege von der Fabrik bis nach Hause die Einkäufe selbst besorgen, so daß es dann meistens schon 16 Uhr ist ehe ich im Heim bin. Hier angekommen, geht es sofort hurtig weiter. Das Mittagessen ist zu kochen; denn der Ehemann und die Kinder haben auch bald Feierabend. In der Zeit von 18 bis 19 Uhr halten wir unseren Mittagstisch. Nachdem und während der Zeit, wo das Mittagessen kocht, werden andere häusliche Arbeiten verrichtet. Bei der Nachmittagsschicht kommt noch hinzu, daß ich an drei Vormittagen neben der

häuslichen Arbeit die Wäsche für die Familie waschen muß. Wir wohnen in einer Siedlung und haben noch zwei Gärten; die erfordern ebenfalls vom Frühjahr bis zum Herbst nicht wenig Arbeit. Von meinem Mann kann ich dabei sehr wenig unterstützt werden. Er verrichtet in einer Hutfabrik schwere Arbeit und ist abends sehr abgespannt."[14]

Not too much had changed when the Women's Office of the Deutsche Arbeitsfront, the Nazi surrogate for a union, made a similar investigation in 1936, the results of which were to be kept strictly secret (no wonder after their promises that women should have the right to be employed but would not need to go out to work for mere survival of their families[15]). Almost a third of the respondents, blue and white collar workers, declared not to have any "Freizeit", and very few had three hours during the week and on Sunday. "Ich habe durchschnittlich 16 Stunden am Tage Arbeitszeit. Ich habe keine Lust und kein Interesse für irgend eine Veranstaltung (they were asked to join the leisure time activities organized by the party and regularly declined). Dazu bin ich viel zu müde. Auch der Sonntag bringt mir keine Abwechslung, da muß ich die Wohnung sauber machen, die Kleidung der Kinder und des Mannes überholen. Die einzige Abwechslung, die ich seit 15 Jahren gehabt habe, war die Geburt eines Kindes und eine 14tägige Krankheit. Ich habe garnichts von meinem Leben. Nun habe ich mich aber daran gewöhnt."

## III.

After a decade of historical investigation of women's activities in and around the home, some of the initially underlying questions have changed: instead of concentrating on proving "that housework was better or more degrading than other kinds of labor done by men", historians now ask what light these investigations may cast "on mentalities, social behavior, and political questions". Such historical research is, however, much older than it may seem today: feminist and female scholars, including historians[16], in the 19th and the early 20th century have dealt with women's domestic work and they often demanded that its value be included in the gross national product. Käthe Schirmacher, a merchant's daughter who preferred to live not with a man but with women and who was one of the earliest German women to receive a doctorate (in 1895), soon afterwards began her work on *Die Frauenarbeit im Hause, ihre ökonomische, rechtliche und soziale Wertung*: "Wenn die Nationalökonomie von 'Frauenarbeit' spricht, so versteht sie darunter fast ausschließlich die Fabrik- und Werkstättenarbeit der Frauen. Die Frauenarbeit im Hause wird meist mit einer kurzen Analyse abgetan, die den nicht produktiven Charakter der häuslichen Frauenarbeit betont. Die Frau im Hause, heißt es, konsumiert Werte, verteilt Werte, schafft aber keine Werte. Ich beeile mich hinzuzusetzen, daß letzteres unrichtig ist." She described at length women's work

in the household, ("Sie sind, je nachdem, Dienstmädchen und Köchin, Wirtschafterin, Schneiderin, Wäscherin, Tapezierer, Maler, Dekorateur, etc."), for the husband ("Die häusliche Frauenarbeit ist die *conditio sine qua non* der außerhäuslichen Berufsarbeit des Mannes") and for the children ("Gibt es endlich eine 'produktivere Arbeit' als die der Mutter? Ist es nicht die Mutter, die ganz allein den Wert aller Werte, den denkenden und handelnden Wert aufbaut, den man ein Menschenwesen nennt?"). She insisted on calling it "real work" even though "it may look like nothing", and added: "Ich muß gegen diese Ausbeutung der Hausfrau und Mutter protestieren, sie ist ebenso ungerecht, verhängnisvoll und unmoralisch wie die der Arbeiterin." She questioned the justice of women being supported not by their own, but by male income and the appearence "daß der Mann für zwei arbeitet, während er doch nur für zwei einstreicht". Like many others among the more radical minority of the women's movement in Western countries around the turn of the century, she demanded "eine Umwertung bestehender Werte" and, more precisely, payment for housework in order to end women's economic and sexual dependency on men.[17]

Helene Stöcker, a single woman, admirer of (a reinterpreted) Nietzsche, advocate of "free love" and among those who called themselves "the radicals" within the women's movement, and Henriette Fürth, a Jewish woman, economist, mother of eight children and advocate of a marriage reform (so that the woman "hört auf, Haussklavin zu sein, und wird dafür zur Herrin. . . Sie hört auf, nur Gebärerin zu sein, und wird dafür zur wirklichen Mutter"), joined in the demand for the acknowledgment of the ideal and economic value of housework. Fürth underlined that this would also lead to improvements in the wages and work conditions of women employed in domestic service, which was then the numerically most important women's job. She argued for a specifically historic consideration of the changes in housework over time: "Es ist eigentümlich, wie wenig wir von uns selbst wissen, von dem Weg, der zu uns hinführt. . . Dieses Nichtwissen ist daran schuld, daß Zeit- für Ewigkeitswerte, vorübergehende Erscheinungen für Dauerzustände gehalten werden." As some historians today, she saw housework not as the "traditional" role of women, but as a product of the 18th and 19th centuries: "Die Beschränkung des Tätigkeitsgebietes der Hausfrau auf die reine Hauswirtschaft und daneben auf die Repräsentation des Hauses, das ist also auf die Dreiheit von Kinder, Küche und Konversation, setzte erst in dem Augenblicke ein, in dem mit den Umwälzungen innerhalb der Produktionstechnik, oder wie es uns geläufiger ist, mit dem Siegeszug der Maschine die Großindustrie aufkam. . . Auch der immer wieder proklamierte und fast zum Axiom erhobene Satz, daß die Frau ins Haus gehöre, ist erst für eine vergleichsweise neue und heute schon wieder in weitem Umfang der Vergangenheit angehörende Zeit zuständig."[18]

Likewise, Marianne Weber, wife of the better-known Max, commented, in 1912, on the still current "slogan" of "wages for housework". She concluded, however, that not its economic, but its "ideal" or "cultural" reevaluation was to be approved of, especially "since every instinct and consciousness of the specific significance of the domestic profession rebels at the material measuring of achievements which cannot be considered market commodities and which, being 'labor of love', are literally immeasurable."[19] During the 1920s, when the "slogan" had lost its attraction, the need for at least an "ideal" reevaluation of housework was often stressed by women, including those who put now their hopes for relief from domestic drudgery in the new possibilities of rationalizing and mechanizing housework. One of them came to the conclusion that "the strongest obstacle to the technological development in the household is the lack of value of women's domestic labor, since it makes the introduction of labor-saving machinery appear unprofitable."[20]

Female housework tended to expand during the economic crisis of the early 1930s, in relation to the period before as well as in relation to the life of unemployed men, as has been shown by the sociologist Marie Jahoda[21] in a study of a town where both men and women had been employed as textile workers. In 1931, such men suffered from too much time, from doing nothing, hanging out in the streets, while the women suffered from overwork for their families, who had to live on less income (e.g. poor relief) than before: "Die Frauen sind nur verdienstlos, nicht arbeitslos im strengsten Wortsinn geworden. . . Der Tag ist für die Frauen von Arbeit erfüllt: Sie kochen und scheuern, sie flicken und versorgen die Kinder, sie rechnen und überlegen und haben nur wenig Muße neben ihrer Hausarbeit, die in dieser Zeit eingeschränkter Unterhaltsmittel doppelt schwierig ist. . . Doppelt verläuft die Zeit in Marienthal, anders den Frauen und anders den Männern. . . So grundverschieden ist die Zeitverwendung bei Männern und Frauen, daß man für sie nicht einmal dieselben Kategorien aufstellen konnte." In the history of working women's lives, not only the experience of work but also the experience of time – of history itself – may be different from that of men, and it requires new and more complex historiographical categories than the traditional ones.

## Notes

[1] Anne Firor Scott, "Woman's Place is in the History Books"(1979), repr. in: id., *Making the Invisible Woman Visible*, Urbana, Ill., 1984, pp. 361ff. On the subject of the invisibility of women's activities see, e.g., id., "On Seeing and Not Seeing; A Case of Historial Invisibility", in: *Journal of American History* 71 (1984), pp. 7-21.

[2] Cf. Marylin Hoder-Salmon, "Collecting Scholars' Wives", in: *Feminist Studies* 4/3 (1978), pp. 107-114.

[3] See, e.g., Nona Glazer-Malbin, "Housework: A Review Essay", in: *SIGNS. Journal of Women in Culture and Society* 1 (1976), pp. 905-922; Ann Oakley, *Woman's Work: The*

*Housewife, Past and Present*, New York 1974; Gerda Lerner, *The Majority Finds Its Past: Placing Women in History*, New York/Oxford 1979, esp. pp. 129-144 ("Just a Housewife"); Gisela Bock/Barbara Duden, Arbeit aus Liebe – Liebe als Arbeit: Zur Entstehung der Hausarbeit im Kapitalismus, in: *Frauen und Wissenschaft*, Berlin 1977, pp. 118-199. A related conceptual change was to conceive gender (gender systems, gender relations, the relation between the sexes, sexual stratification) not as a "biological", but as a social category: cf. Joan Kelly, "The Social Relations of the Sexes: Methodological Implications of Women's History", in: *SIGNS* 1 (1976), pp. 809-824; Susan Carol Rogers, "Woman's Place: A Critical Review of Anthropological Theory", in: *Comparative Studies in Society and History* 20 (1978), pp. 123-162.

[4] International Labor Organization data, quoted in: Hilda Scott, *Working Your Way To The Bottom: The Feminization of Poverty*, London etc. 1984, p. 3.

[5] Lorenz von Stein, *Die Frau auf dem Gebiete der Nationalökonomie*, Stuttgart(6) 1886, pp. 98-99, 111; for this use of "Mensch" in contrast to "Frau" cf., e.g., p. 137. The previous quotes: pp. V, 7-8, 144, 48-49; the following quotes: pp. 94-95, 112-114, 84-85, 67, 117.

[6] This was the formula used by an expert on poor relief in 1889, quoted in: Gerda Thornieport, *Studien zur Frauenbildung*, Weinheim/Basel 1979, p. 131. Cf. Gustav Schmoller, *Grundriß der Allgemeinen Volkswirtschaftslehre* (1900), Leipzig 1901, pp. 249-252.

[7] *Handwörterbuch der Staatswissenschaften*, vol. 5, Jena 1923, p. 161, art. "Haushaltung".

[8] Eugene D. Genovese, *Roll Jordan Roll. The World the Slaves Made*, New York 1974, pp. XXIf.; for the scholarly work of this scholar's wife see Elizabeth Fox-Genovese, "Placing Women's History in History", in: *New Left Review* 133 (1982); another important overview is Olwen Hufton/Joan W. Scott, "Women in History", in: *Past and Present* 101 (1983), pp. 125-157.

[9] For the latter cf. Lyndal Roper, "Housework and Livelihood", in: *German History* 2 (1985), pp. 3-9, who explores this subject for 16th-century Augsburg and deals with some of the older and newer cultural definitions of housework.

[10] Sibylle Meyer, "Die mühsame Arbeit des demonstrativen Müßiggangs: Ueber die häuslichen Pflichten der Beamtenfrauen im Kaiserreich", in: *Frauen suchen ihre Geschichte. Historische Studien zum 19. und 20. Jahrhundert*, ed. Karin Hausen, Munich 1983, p. 172; cf. id., *Das Theater mit der Hausarbeit: Bürgerliche Repräsentation in der Familie der wilhelminischen Zeit*, Frankfurt a.M./New York 1982; for a similar view on British middle-class women see Patricia Branca, "Image and Reality: The Myth of the Idle Victorian Woman", in: *Clio's Consciousness Raised: New Perspectives on the History of Women*, ed. Mary S. Hartman/Lois Banner, New York etc. 1974, pp. 179-191.

[11] Ruth Schwartz Cowan, More Work for Mother: The Ironies of Household Technology from the Open Hearth to the Microwave, New York 1986; Susan Strasser, *Never Done: A History of American Housework*, New York 1983; Joann Vanek, "Time Spent in Housework", in: *Scientific American* (Nov. 1974), pp. 116-120.

[12] Heidi I. Hartmann, The Family as the Locus of Gender, Class, and Political Struggle: The Example of Housework, in: *SIGNS* 6/3 (1981), pp. 366-394.

[13] E.g. Bock/Duden (note 3).

[14] *Mein Arbeitstag – Mein Wochenende: 150 Berichte von Textilarbeiterinnen*, ed. by Arbeiterinnensekretariat des Deutschen Textilarbeiterverbands, Berlin 1930, pp. 145, 224.

[15] This was the content of the nazi propaganda campaign in relation to women for the election of November 1932, when for the first time a larger number of women voted for the NSDAP (Thomas Childers, *The Nazi Voter*, Chapel Hill/London 1983, pp. 188f., 239f). The results of the Frauenamt der DAF: Bundesarchiv Koblenz, NS 5 I/3-4. Cf. Dörte Winkler, *Frauenarbeit im "Dritten Reich"*, Hamburg 1977.

[16] See Bonnie G. Smith, "The Contribution of Women to Modern Historiography in Great

Britain, France, and the United States, 1750-1940", in: *American Historical Review* 89 (1984), pp. 709-752; the previous quote is from p. 730; on p. 715 the author rightly points out that women historians today have been less fascinated than their predecessors with women's intellectual work. Cf. also Kathryn Kish Sklar, "American Female Historians in Context, 1770-1930", in: Feminist Studies 3/1-2 (1975/6), pp. 171-184; Natalie Zemon Davis, "Gender and Genre: Women as Historical Writers, 1400-1820", in: *Beyond Their Sex: Learned Women of the European Past*, ed. Patricia Labalme, New York 1980, pp. 153-182.

[17] Käthe Schirmacher, *Die Frauenarbeit im Hause, ihre ökonomische, rechtliche und soziale Wertung* (1905), Leipzig 1912, pp. 3-8, 11; for the "radicals" see, e.g., *Feministische Studien* 3/1 (1984): "Die Radikalen der alten Frauenbewegung".

[18] Henriette Fürth, *Die Hausfrau*, München 1914, pp. 35-42; for comparable hypotheses see Bock/Duden (note 3) and Thornieport (note 6). For Helene Stöcker see *Feministische Studien* (note 17).

[19] Marianne Weber, Zur Frage der Bewertung der Hausfrauenarbeit (1912), in: id., *Frauenfragen und Frauengedanken*, Tübingen 1919, p. 89.

[20] Erna Meyer, *Der neue Haushalt*, Stuttgart 1926, p. 294.

[21] Marie Jahoda/Paul F. Lazarsfeld/Hans Zeisel, *Die Arbeitslosen von Marienthal*, Leipzig 1933, repr. Allensbach/Bonn 1960 and Frankfurt a.M. 1975. The investigation of the lives of the Marienthal women was done by Marie Jahoda and other women. The following quotes: ed. 1975, pp. 84, 89f.

# L'industrie électrotechnique allemande entre les deux guerres: à la recherche d'une position internationale perdue

PETER HERTNER
Damstadt/Florence

A la différence de l'industrie chimique allemande qui, en 1929, avait réussi à reconquérir sa part du marché mondial de l'avant-guerre[1], l'industrie électrotechnique allemande, après 1918, enregistra un recul permanent même si, de façon relative, vue sa participation à l'exportation mondiale, elle se trouvait encore en tête devant les États-Unis et la Grande-Bretagne. Avant 1914, elle détenait presque la moitié de l'exportation mondiale; à partir du milieu des années 20, sa participation, en règle générale, sera d'un bon quart, elle n'approchera le tiers que lors des années de crise de 1930 à 1933 (voir tableau en annexe). Cet état de choses était bien connu des observateurs de l'époque. Ainsi, la British Electrical and Allied Manufacturers' Association, lors d'une analyse des rapports de force sur le marché électrotechnique en Europe en 1927, établit-elle[2]:

> "Before the war the great manufacturing firms of Germany and the United States carried out a wholly successful policy of financial penetration into overseas markets through the formation of power finance corporations which raised the capital required for new power schemes, created, in a number of cases, special supply undertakings, and acted as technical and financial consultants. There is no doubt that before the war a high percentage of German electrical exports to South America and European countries generally was due to direct and indirect financial connections... The war and its aftermath destroyed much of the system so carefully prepared by Germany, and allowed Switzerland and Belgium to obtain a hold in many countries once considered a German preserve, such as Italy, Spain, Poland and South America. The balance of strength in power finance has shifted to these two countries, while the United States are now carrying out an aggressive policy."

Je reviendrai plus tard aux problèmes de financement, je tiens d'abord à présenter à quelle échelle se déployait la concurrence au niveau du marché mondial: en 1912, l'Allemagne produisit du matériel électrotechnique pour une valeur totale de 1 200 millions de marks, dont 22% furent exportés. La production de la Grande-Bretagne fut seulement de 450 millions, mais 29% furent exportés. Les États-Unis enfin eurent en 1912 un volume de production correspondant à 1 500 millions de marks, dont toutefois 7%

seulement furent exportés, tout le reste étant absorbé par le grand marché intérieur[3]. Après la guerre, l'écart entre le volume de production allemand et américain avait considérablement augmenté: en 1927, on produisit aux États-Unis du matériel électrotechnique pour une valeur de 7 006 millions de marks, en Allemagne pour une valeur de 2 700 millions; certes, la production allemande d'avant-guerre avait ainsi plus que doublé, mais son taux de progression était de loin inférieur à celui de l'Amérique. En cette même année 1927, la Grande-Bretagne et la France suivaient avec un volume de production correspondant respectivement à 1 700 et à 650 millions de marks. Sur le marché mondial relativement stationnaire de l'entre-deux-guerres, le taux d'exportation avait également diminué; en 1927, il n'était plus que de 16% pour l'industrie électrotechnique allemande; le taux d'exportation des USA, d'environ 6%, n'avait pas notablement changé[4].

Depuis la fin de la crise économique des années 1901-1902, le marché allemand était dominé par deux grandes entreprises, le konzern *Siemens* et le konzern de l'*Allgemeine Elektrizitäts-Gesellschaft* (AEG). Déjà en 1902, d'après une estimation de l'époque, ils détenaient, avec les fabricants de câbles *Felten & Guilleaume*, environ les trois quarts de la production électrotechnique allemande[5] et, selon une estimation ultérieure contemporaine pour 1918, la situation était jusque-là restée inchangée[6]. Pour l'année 1936, un calcul précis démontre que les deux grands konzerns *Siemens* et *AEG* réunissaient 51% de la production électrotechnique totale[7]. D'un point de vue international, les deux firmes allemandes avaient, à la fin des années 30, un chiffre des ventes de loin inférieur à celui de leurs concurrents américains mais pouvaient encore se prévaloir d'une importance respectable, et du moins en Europe étaient largement en tête. En 1929, *Siemens* avait un chiffre d'affaires de 850 millions de marks et *AEG* de 580 millions; la *General Electric* et la *Western Electric*, la société de fabrication de l'*American Telephone and Telegraph Company* (ATT) avec un chiffre d'affaires respectivement de 1 744 et de 1 722 millions de marks, les distançaient certes considérablement, mais le deuxième producteur de matériel lectrique à courant fort aux États-Unis, la *Westinghouse Electric and Manufacturing Company*, avec un chiffre d'affaires de 909 millions de marks, n'était déjà plus tellement éloigné du konzern *Siemens*[8].

Les deux grandes entreprises allemandes se différencient considérablement l'une de l'autre quant à leur évolution historique et à leur structure de production telle qu'elle s'était formée jusqu'à la fin des années 20. La *Siemens & Halske*, fondée à Berlin en 1847, concentra très fortement son activité, jusqu'à une période avancée des années 1880, sur le secteur des courants faibles, en particulier sur le domaine télégraphique. Ce fut en outre – et totalement après le retrait du partenaire Halske en 1867 – une entreprise familiale, dominée par la personnalité de Werner Siemens, inventeur et entrepreneur, qui dans un but bien précis fit participer ses

frères cadets à la direction de l'entreprise[9]. L'*AEG*, par contre, avait été fondée en 1883 par Emil Rathenau, sous le nom de *Deutsche Edison-Gesellschaft für angewandte Elektrizität*, grâce à l'aide décisive d'un groupe bancaire berlinois et avec le concours de la *Siemens & Halske*. Dès le début, elle exerça une activité propre dans le secteur des courants forts, tout d'abord limitée à la production et à la distribution d'électricité à Berlin. Les étroites relations techniques avec Siemens, qui, au début, ne permettaient à la société de Rathenau que de produire des ampoules, furent en deux étapes, en 1887 et 1894, complètement dissoutes; en 1887, la *Deutsche Edison-Gesellschaft* fut rebaptisée *Allgemeine Elektrizitäts Gesellschaft*[10]. Lorsque *Siemens*, bien qu'avec du retard, se tourna davantage, au cours des années 1890, vers le secteur des courants forts, la situation concurrentielle s'accentua entre cette entreprise familiale, transformée en 1897 en société par actions, et la "managerial enterprise" *AEG* dirigée par Emil Rathenau[11]. Les deux firmes surmontèrent relativement bien la crise de 1901-1902 et eurent même le rôle principal dans le processus de concentration sans pareil qui par suite de la crise se déclara dans l'industrie électrotechnique allemande[12]. En 1903, la *Siemens & Halske* reprit la firme *Schuckert* de Nuremberg, troisième firme électrotechnique en Allemagne, qui en 1899-1900, avec un chiffre d'affaires total de 77 millions de marks, avait pratiquement rejoint le chiffre d'affaires réalisé par *Siemens* sur le marché intérieur, mais qui par la suite se trouva confrontée à une sévère crise de liquidité[13]. L'activité de la *Schuckert* s'était exercée exclusivement dans le domaine des courants forts; ainsi les usines de courants forts de la *Siemens* furent-elles réunies à la *Schuckert* lors de la fusion de 1903 sous le nom de *Siemens-Schuckert-Werke*. La *Siemens & Halske* ne fut dès lors plus responsable que du secteur des courants faibles[14].

Cette même année 1903, l'*AEG* reprit également l'*Union-Elektrizitäts-Gezellschaft*, qui était étroitement liée respectivement à la *Thomson-Houston* d'abord et ensuite à la *General Electric*. En même temps, l'*AEG* conclut avec la *General Electric* un accord prévoyant l'échange mutuel des brevets et d'expériences ainsi que la "délimitation des zones d'activité dans le monde entier". En fonction de cet accord, la *General Electric* devait obtenir comme "zone exclusive" l'Amérique du Nord, l'*AEG* l'Europe centrale, la Scandinavie, l'Europe de l'Est et du Sud-Est. En Grande-Bretagne et en France était prévue une étroite collaboration, en Italie la fondation d'une société commune. Pour le reste du monde, on envisageait une entente ultérieure[15].

C'est ainsi qu'en Allemagne fut pratiquement créé un duopole composé des donzerns *Siemens* et *AEG*, dont la position dominante sur le marché se trouva encore renforcée par le fait que les seules moyennes entreprises de l'électrotechnique encore existantes durent le rallier: ainsi l'*AEG* reprit-elle en 1910 les *Lahmeyer-Werke*[16], et, en 1912, *Siemens*, avec l'aide de la *Deutsche Bank* qui depuis les années 1890 lui était particulièrement liée,

réussit à obtenir, moyennant un engagement de capitaux relativement réduit, le contrôle des *Bergmann-Elektrizitätswerke*[17]. Durant la dernière décennie avant la Première Guerre mondiale, la croissance des deux grands konzerns fut particulièrement impressionnante: le chiffre d'affaires réuni de la *Siemens & Halske* et des *Siemens-Schuckertwerke* passa de 92 millions de marks au cours de l'exercice 1903-1904 à 414 millions de marks en 1913-1914. Le nombre des employés du konzern *Siemens* en Allemagne et à l'étranger passa de 25 800 en 1903 à 81 800 en 1913[18]. A l'*AEG*, l'effectif, au cours de cette même période, passa de 18 300 à 68 700 personnes[19].

Lorsqu'éclata la Première Guerre mondiale, chacune des deux grandes entreprises détenait à l'étranger des intérêts considérables: *Siemens* possédait des établissements de production en Autriche, en Hongrie, en Russie, en Espagne, en Grande-Bretagne et en France[20]. En 1914, les investissements dans ces succursales à l'étranger s'élevaient à 115 millions de marks, ce qui correspondait à 19,5% du capital fixe et du capital d'exploitation inscrits au bilan de tout le konzern[21]. A cette même époque, 24 606 dépendants de Siemens, c'est-à-dire 30% de l'effectif total, étaient employés dans les succursales à l'étranger et dans les 168 autres établissements et bureaux techniques également à l'étranger[22]. Alors que Siemens avait déjà fondé, en 1855 à Saint-Pétersbourg et en 1858 à Londres, des succursales auxquelles étaient rattachés des établissements de production fonctionnant de façon encore entièrement artisanale[23], l'*AEG* ne put se décider à créer des établissements à l'étranger qu'au début du siècle nouveau. L'occasion lui en fut fournie par la fusion susmentionnée avec l'*Union-Elektrizitäts-Gesellschaft* en 1903; l'*AEG* reprit les usines de cette dernière à Riga et à Vienne et fonda en 1904, avec la *Compagnie Thomson--Houston de la Méditerranée*, une société à Milan qui, à partir de 1907, se lança également dans la production électrotechnique[24]. Le motif de la création d'une usine à Milan, lequel pourrait tout aussi bien s'appliquer aux filiales de production de *Siemens* et d'*AEG* en Russie et en Autriche--Hongrie, a été décrit par le consul allemand à Rome dans son rapport commercial pour l'année 1910:

> "Les grandes sociétés comme l'AEG et la Brown Boveri ont créé des succursales grâce auxquelles non seulement elles économisent des droits de douane mais encore peuvent tranquilliser les protectionnistes du pays du fait qu'elles présentent des produits nationaux"[25].

Vers le milieu de 1912, 5 400 personnes, c'est-à-dire 8% de l'effectif de l'*AEG*, étaient employées dans ses usines à l'étranger[26].

La position dominante de l'industrie électrotechnique allemande sur le marché mondial, telle qu'elle a déjà été exposée ci-dessus, le fait qu'en 1913, à titre d'exemple, 87% des importations russes et 70% des importations italiennes de matériel électrotechnique provenaient d'Allemagne, ainsi que le fait, par exemple, qu'entre 1903 et 1913 environ la moitié des

importations argentines de produits électrotechniques venait d'Allemagne[27], tout cela peut s'expliquer par trois causes essentielles:

1. La *capacité technique élevée*, du moins par rapport aux autres États industriels européens[28], des producteurs électrotechniques allemands qui, grâce à une rationalisation étudiée et à une standardisation de la production, purent passer de bonne heure à une fabrication de masse avantageuse, comme ce fut par exemple le cas de l'*AEG* à partir de 1897 dans le domaine de la production de petits moteurs électriques[29].

2. Un autre facteur fut sans aucun doute la *multinationalisation croissante* des entreprises au moyen de l'élargissement mondial, dont nous avons déjà parlé, d'un réseau de distribution grâce à la création de représentations, de succursales de vente et de bureaux techniques. De cette façon, grâce à une internationalisation croissante, les coûts de transaction purent être diminués[30]. Le dernier pas vers l'internationalisation fut ensuite la création d'unités de fabrication dans le pays d'accueil. Si l'avance relative des producteurs allemands au niveau du prix et de la qualité peut être qualifiée, dans la conception développée par John H. Dunning, de "ownership specific endowment", la politique douanière et les mesures non tarifaires des pays de destination (comme par exemple la préférence accordée aux producteurs nationaux en cas de commandes d'État) peuvent être appelés "location-specific"[31]. Felix Deutsch, successeur des deux Rathenau à la tête de l'*AEG*, a clairement exprimé la situation de contrainte dans laquelle se trouvaient les exportateurs électrotechniques en face du protectionnisme grandissant: en particulier vis-à-vis des mesures non tarifaires, il n'y avait "que deux alternatives", "... ou renoncer complètement aux affaires, ou mettre sur pied une certaine fabrication dans le pays même ... C'est pourquoi, dans certains pays, on se décida à faire soi-même certaines pièces, rendues inutilement plus chères par le transport et la douane". Lorsque Deutsch constate finalement, "... nous avons toujours considéré nos usines à l'étranger uniquement comme – dirais-je – des établissements de sous-traitance pour nos usines allemandes"[32], on ne doit pourtant pas avancer cet argument "défensif" contre une tendance progressive tout à fait réelle à la multinationalisation[33]. Ce sont justement de telles déclarations de l'époque qui éclairent tout particulièrement le rapport étroit entre exportation et investissement direct ainsi que l'obligation d'une internationalisation progressive malgré d'importants intérêts à la production dans la métropole.

3. Enfin, ce qu'on appelle *Unternehmergeschäft*, dans le cadre duquel les grandes entreprises électrotechniques, outre la création de centrales électriques, de tramways électriques et d'installations d'éclairage, s'occupaient également de leur financement, a sûrement joué un rôle important dans la pénétration des marchés étrangers. En fin de compte, cela se fit uniquement pour soutenir la vente des propres produits électrotechniques, c'est-à-dire pour créer un marché par le vendeur lui-même, et ceci en

particulier dans les pays du sud et de l'est de l'Europe ainsi qu'en Amérique latine, où manquaient le capital à risques et l'initiative d'entreprise.

L'*Unternehmergeschäft*, tel qu'il était conçu au début, conduisit bien sûr rapidement à une charge financière extravagante pour les producteurs électrotechniques avec des installations en cours de construction, et même après l'achèvement de celles-ci, le placement des actions et des obligations des centrales électriques, des tramways et des sociétés d'éclairage se transforma le plus souvent en une opération de longue haleine, qui pouvait représenter une menace sérieuse pour la liquidité des sociétés mères. La création de sociétés de financement spéciales devait remédier à cet état de choses. Celles-ci pratiquaient la "substitution de titres", en ce sens qu'elles reprenaient dans leur portefeuille les titres (actions et obligations) des sociétés d'infrastructure nouvellement fondées, et émettaient leurs propres titres à la place de ceux-ci. Dès que les sociétés d'infrastructure avaient achevé leurs installations, ce qui, en particulier pour les investissements hydro-électriques, pouvait demander plusieurs années, et dès que ces installations commençaient à rapporter un bénéfice, les sociétés de financement se mettaient peu à peu à négocier sur le marché des capitaux les titres tenus dans leur portefeuille, mais gardaient en règle générale une participation minoritaire de contrôle[34]. Les fondateurs de ces financières étaient les konzerns électrotechniques ainsi qu'une série de grandes banques. Chacun des konzerns donna le jour non à une mais à plusieurs financières; leurs sièges sociaux furent élus la plupart du temps en Suisse ou en Belgique, car ces deux pays offraient des avantages fiscaux et la plus grande libéralité possible sur le plan du droit commercial[35].

Dès le début, les sociétés financières furent tenues par des stipulations à recevoir le matériel électrotechnique nécessaire à l'édification de leurs centrales électriques, tramways et sociétés d'éclairage, exclusivement de leur propre "konzern mère"[36] – ceci représentant également un acte conscient d'internationalisation qui, d'une part, coupait du marché les nouvelles sociétés d'infrastructure et, d'autre part, comme on l'a déjà laissé entendre, assurait aux konzerns électriques un marché garanti.

La plupart des financières furent fondées au cours de la deuxième moitié des années 1890, et non pas seulement par *Siemens* et *AEG*, mais également par de plus petites firmes; je citerai ici trois d'entre elles, représentatives de toute une série:

a) L'une des plus importantes était sans aucun doute la *Banque pour entreprises électriques*, en allemand abrégée le plus souvent en *Elektrobank*, fondée en 1895 à Zürich par l'*AEG*, la *Deutsche Bank*, la *Berliner Handelsgesellschaft*, le *Crédit suisse* ainsi que par d'autres banques allemandes, suisses, françaises et italiennes. Très vite après sa formation, elle s'occupa des centrales électriques et des sociétés de tramways créées par l'*AEG* à Gênes et dans diverses villes espagnoles et finança au cours des deux

décennies suivantes un grand nombre d'autres créations de l'*AEG* en Allemagne, en Europe du Sud et de l'Est[37].

b) Une autre importante financière était la *Société suisse d'industrie électrique*, en abrégé nommée le plus souvent "Indelec"[38], fondée en 1896 à Bâle par *Siemens & Halske* en association avec un consortium de banques suisses et allemandes. Contrairement à l'*AEG* qui détint pendant un certain temps 95% du capital de l'*Elektrobank*, *Siemens* ne disposait à l'*Indelec* que d'une participation minoritaire et voyait à l'occasion mise en question son influence décisive auprès de cette financière[39]. L'*Indelec* également finança jusqu'à la Première Guerre mondiale toute une série de sociétés d'infrastructure, le plus souvent liées au konzern *Siemens*, en Suisse, en Allemagne, en Russie, en France et en Italie. Sa puissance financière n'atteignait pas celle de l'*Elektrobank*, ce qui peut être dû au fait que les banques fondatrices de l'*Indelec*, comparées à celles de l'*Elektrobank*, n'étaient que de second plan[40].

c) La *Société financière de transports et d'entreprises industrielles* (Sofina), fondée à Bruxelles en 1898, ne se laisse pas aussi clairement rattacher à un konzern particulier. A sa fondation et à sa première augmentation de capital (1899) participa de façon majoritaire le capital allemand, c'est-à-dire aux côtés de la *Dresdner Bank* et de la *Disconto-Gesellschaft*, surtout la *Gesellschaft für elektrische Unternehmungen* (Gesfürel), qui de son côté faisait aussi office de financière pour l'*Union-Elektrizitätsgesellschaft*. La *Thomson-Houston* et respectivement la *General Electric* étaient ainsi, également d'une manière indirecte, engagées dans la Sofina. La fusion en 1903 de l'*Union* et de l'*AEG* introduisit la Sofina dans l'empire de Rathenau, et par la suite la Sofina finança toute une série de sociétés affiliées au konzern *AEG* en Europe occidentale et en Argentine, mais en outre également un certain nombre d'entreprises auxquelles était intéressé surtout le capital belge[41]. Les relations avec la General Electric semblent avoir subsisté et, dans l'ensemble, on pourra dire que, des trois cas cités, la Sofina paraît être la plus apte à personnifier l'internationalisation du capital avant 1914.

La défaite de l'Allemagne lors de la Première Guerre mondiale modifia radicalement la position internationale de l'industrie électrotechnique allemande ainsi que son ingénieux système de vente et de participation. De l'ensemble des investissements allemands à l'étranger, évalués pour 1913 autour de 20 à 25 milliards de marks, environ 15 milliards furent perdus du fait du séquestre dans les États ennemis et de la vente pendant la guerre[42]. Selon une autre estimation de l'époque, des 11 milliards de marks des investissements directs allemands, tels qu'ils se présentaient avant la guerre, restait à la fin de 1922 1,8 milliard[43]. D'après une enquête effectuée en 1930, les investissements allemands à l'étranger étaient alors de 8,8 à 10,8 milliards de reichsmarks. Ils avaient pratiquement doublé depuis 1924 – cela toutefois avec une dette extérieure simultanée de 26,1 à 27,1 milliards de reichsmarks. On peut estimer les investissements

directs allemands en 1930 à environ 2-2,5 milliards de reichsmarks[44].

Toujours est-il qu'en 1919 *Siemens* et *AEG* durent déplorer la perte de beaucoup de leurs succursales à l'étranger. En ce qui concerne les usines, seuls purent être sauvés les établissements se trouvant en Autriche, en Hongrie et en Espagne, laquelle était restée neutre. La poursuite de l'*Unternehmergeschäft*, qui avait eu tant de succès dans l'avant-guerre, fut radicalement mise en question par la perte du contrôle sur les financières dont le siège se trouvait en Suisse et en Belgique. Dans le cas de la Belgique, cela se produisit assez tôt: "Pendant la guerre, la Sofina se détachait du groupe allemand et commençait à s'appuyer fortement sur des établissements financiers belges..."[45]. Dans le cas de la Suisse, l'*Elektrobank* et l'*Indelec* enregistrèrent de graves pertes de revenus du fait de la guerre et de l'inflation de l'après-guerre. Pour l'apport de nouveaux capitaux pour les deux financières, ni l'*AEG*, ni *Siemens*, ni les banques allemandes ne purent réunir les fonds nécessaires, de telle sorte que leur participation au capital se réduisit dans chacun des deux cas à bien moins de 25%[46].

Dans cette situation, de quels moyens disposaient les deux grandes entreprises pour lutter contre la menace de la perte des marchés étrangers? Je voudrais en énumérer brièvement quelques-uns.

1. Après la perte de la plupart des usines à l'étranger, il fallut reconstituer au plus vite le réseau de succursales commerciales et de bureaux techniques, afin de pouvoir exporter avec succès. Les deux sociétés s'y employèrent dès la fin de la guerre[47]. Pendant les premières années, le déclin rapide du mark aida également à regagner des positions dans l'exportation. En profita surtout le secteur de haute tension, dans lequel l'industrie allemande ne semble pas avoir eu une avance technologique particulière. A la fin de 1923, par exemple, *Siemens-Schuckert* avait de nouveau atteint dans l'exportation vers les pays extra-européens le volume réel des commandes de 1913[48].

2. A cause de la pénurie de capitaux, les investissements directs dans les unités de fabrication à l'étranger devaient être limités à un minimum. Dans divers cas, ils étaient cependant inévitables si on ne voulait pas perdre un marché précis. Comme dans l'avant-guerre, nombreux furent les pays qui, pour les commandes du secteur public, donnèrent la préférence aux "producteurs nationaux". Ceci s'observa surtout dans les États successeurs de la monarchie des Habsbourg, et c'est ainsi que *Siemens*, en Tchécoslovaquie par exemple, dut créer à partir de 1920 trois sociétés dotées de leurs propres établissements de production, dans le domaine des courants faibles, afin de rester en relation d'affaires avec l'État tchécoslovaque pour des livraisons militaires et les équipements téléphoniques[49]. Tandis qu'à l'origine ces sociétés importaient encore la plus grande partie des produits semi-finis en les faisant fabriquer par les konzerns mères en Allemagne, les restrictions à l'importation, en augmentation rapide depuis le début des années 30, contraignaient à fabriquer de plus en plus presque

toutes les pièces en Tchécoslovaquie même[50]. De même, la filiale de l'*AEG* en Hongrie se chargea au cours des années 30 de la production de divers articles électrotechniques[51], et sa filiale de Prague dut certainement, si l'on considère le volume de son capital, s'occuper également de la production[52]. Même en France, il ne fut possible à *Siemens & Halske* d'être un concurrent avec succès, pour son système téléphonique automatique, qu'après avoir pris une participation en 1927 à la *Compagnie générale de télégraphie et de téléphone*[53]. Au Danemark, en 1935-1936, *Siemens* dut accepter une *joint-venture* avec une firme locale afin de ne pas laisser le secteur téléphonique complètement aux mains de l'*ITT*, et il fallut, pour entrer seulement en ligne de compte lors de la distribution des commandes, présenter au gouvernement danois au moins la perspective d'une production dans le pays[54]. Au Japon également, *Siemens* conclut en 1923, avec un apport de capital de 30%, une *joint-venture* avec une firme japonaise, seul moyen de pénétrer un marché de nos jours encore réputé difficile[55]. En 1929, *Siemens* réussit enfin à obtenir une participation minoritaire de 15% à la *Siemens Brothers*, l'ancienne succursale britannique, séquestrée pendant la guerre et ensuite expropriée. Cela impliquait une étroite collaboration technique et une prise réciproque de brevets et de licences; en même temps, *Siemens & Halske* avait gagné un allié important dans la lutte pour les affaires téléphoniques au niveau mondial contre l'*ITT* américaine[56].

3. Un autre moyen de pouvoir se maintenir en ligne pour la participation au marché mondial, malgré la pénurie de capitaux et la perte de contrôle sur les grandes sociétés financières, fut le recours aux capitaux américains qui, à partir de 1925, arrivèrent de plus en plus en Allemagne et dans l'Europe entière[57]. La participation de 25% prise en 1929 par la *General Electric* dans l'*AEG*, qui se trouvait dans des difficultés de liquidité aiguës[58], est devenue particulièrement célèbre, mais la *General Electric* participa également avec 16 2/3% au capital de la *Osram GmbH*, création commune de *Siemens* et de l'*AEG* pour la production d'ampoules[59]. *Siemens* se tourna également vers le capital américain et prit, à partir de 1925, avec l'aide de la banque new-yorkaise *Dillon Read & Co.*, jusqu'en 1930, plusieurs emprunts en dollars sur le marché de capitaux américain[60].

4. Un autre moyen pour sortir de l'isolement international causé par la guerre fut la coopération technique, c'est-à-dire l'échange d'expériences et de brevets, avec les grandes sociétés américaines, et c'est ainsi qu'en 1923 l'*AEG* et la *General Electric* renouèrent leurs liens traditionnels par un "contrat d'amitié", renouvelé et approfondi en 1929[61]. De son côté, *Siemens* choisit en 1924 comme partenaire l'autre producteur américain important de matériel de haute tension, la société *Westinghouse*, pour conclure un accord technique de dix ans, renouvelé en 1934[62].

5. Restait la possibilité d'"accords" sous forme de cartels, mais le caractère de l'industrie électrotechnique avec ses produits en général assez

"sophistiqués" et hétérogènes ne se prêtait pas très bien à des ententes pareilles. En effet, il y en avait très peu; la plus connue était le cartel des producteurs d'ampoules, fondé en 1924, qui réunissait à peu près 90% de la production mondiale sans les États-Unis et le Japon[63]. A côté de cela, on trouve plutôt des ententes générales de grands producteurs, créées pour se partager les marchés, et en général non limitées à un seul produit. Comme exemple, on pourrait mentionner la "convention de Zürich" pour les concessions téléphoniques, conclue en 1929 par l'*Autelco* américaine, *Siemens Brothers* de Londres et *Siemens & Halske*. C'était au fond une alliance contre l'*ITT* américaine qui, depuis 1925, conquérait un marché après l'autre, grâce à ses stratégies agressives[64]. Dans le secteur de la haute tension, particulièrement sensible à la conjoncture, la grande crise accélérait pourtant les expériences visant une entente assez vaste, et c'est ainsi qu'en 1931 les grands producteurs américains, britanniques, allemands et la *Brown Boveri* suisse se réunissaient pour signer l'*International Notification and Compensation Agreement*, un cartel qui enregistrait, distribuait et fixait les prix des commandes pour les régions extra-européennes et qui fonctionna assez bien jusqu'au début de la Seconde Guerre mondiale[65].

6. Enfin, il y avait un dernier paramètre que les deux grands konzerns allemands pouvaient employer pour améliorer leur situation au niveau du marché mondial. D'une part, ils pouvaient faire baisser leurs coûts de production par un effort énergique de rationalisation et, en effet, *Siemens* et *AEG* s'y employèrent tous deux à partir de 1924, c'est-à-dire après la fin de la grande inflation[66]. D'autre part, il était toujours possible de modifier le programme de production en mettant l'accent sur la fabrication de produits plus rentables et moins sensibles aux variations de la conjoncture. Au cours des années 1930-1933, c'était sans doute les produits du secteur de la haute tension qui étaient les plus touchés, et on peut dire que l'heure de *Siemens* était arrivée, lequel, depuis un demi-siècle au moins, avait systématiquement soigné les deux secteurs et avait même souffert, pendant une période assez longue, de la croissance nettement plus lente du domaine de la faible tension. Au moment de la crise, ce sont surtout les affaires du téléphone qui maintiennent l'entreprise *Siemens*, tandis que la *Siemens-Schuckert*, qui réunit tout le secteur de la haute tension, voit son chiffre d'affaires réduit en 1932-1933 à un tiers de celui de l'année 1928-1929[67]. Ce n'est donc pas du tout surprenant que l'*AEG* ait ressenti la crise de façon bien plus grave, bien qu'elle ait pris une ampleur considérable également dans le secteur de la faible tension depuis les années de la guerre[68]. Les années de la crise et celles qui suivirent favorisèrent donc nettement les entreprises ayant su diversifier leurs programmes de production au maximum.

Au lieu de tirer une conclusion,, on pourrait peut-être mieux illustrer nos thèses en citant les cas de trois pays bien différents quant à leur histoire et à leur situation géographique, c'est-à-dire les cas de l'Italie, de

l'Espagne et de l'Argentine. Si l'on considère l'évolution des affaires de l'entreprise *Siemens* en Italie entre les deux guerres, on remarque bien des éléments mentionnés déjà auparavant: difficultés croissantes sur le plan des ventes de produits de haute tension, une fois passées les années de l'exportation facile pendant la grande inflation allemande – il s'agit évidemment d'un secteur dans lequel une entreprise sans unités de production à l'étranger se trouve relativement désavantagée à cause du protectionnisme (tarifaire et non tarifaire) montant[69]; par contre, succès relatif de l'électrochimie et surtout du secteur de faible tension où *Siemens*, soutenue par des financements américains, réussit à s'implanter solidement dans les concessions téléphoniques[70]; succès aussi dans le secteur des produits de l'électromédecine et de l'armement (fournitures à la marine de guerre italienne). Lorsque l'autarcie fasciste devient plus menaçante, *Siemens* peut "mobiliser" les quelques petites unités de production qu'elle contrôle dans le pays et devenir ainsi de plus en plus un producteur "national"[71].

Le cas de l'Espagne est moins favorable. Poussée par le tarif douanier "ultraprotectionniste" de 1906 vers la production "nationale", *Siemens* y possède depuis 1910 une usine en Catalogne qui prospère pendant la guerre[72]. Les années 20 sont beaucoup plus difficiles, à cause surtout de la concurrence américaine (*General Electric* et *ITT*) de plus en plus agressive. Le projet d'une fusion avec l'*AEG* espagnole ne se réalise pas car les conceptions des deux entreprises divergent très fortement, et ce sera la même chose en Allemagne où une pareille fusion ne se réalise pas non plus[73]. Face à un marché peu dynamique, tous les grands producteurs de matériel de haute tension présents en Espagne, européens et américains, s'associent dans un cartel à la fin des années 20[74].

Enfin l'Argentine. Là, les Allemands – l'*AEG*, *Siemens-Schuckert* et les grandes banques berlinoises – avaient perdu, en 1919, le contrôle de la *Deutsche-Überseeische Elektrizitätsgesellschaft*, fondée en 1898 et qui était devenue, avec un capital nominal de 150 millions de marks en 1914, la plus grande financière électrique en Amérique latine[75]. En 1920, elle fut transformée en la *Compañía Hispano Americana de Electricidad*, son siège fut transféré de Berlin à Barcelone et ce fut alors surtout la Sofina qui y exerça une influence dominante[76]. Plus tard, en 1926-1927, les deux konzerns allemands perdent une bataille très significative en Argentine contre les Américains de la *General Electric* et de sa société financière, l'*Electric Bond and Share Company*. Ce sont les Américains qui l'emportent donc et réussissent à acheter une importante compagnie nationale. Les sociétés berlinoises ont compté en vain sur l'assistance financière de la Sofina, invoquant les liens traditionnels. C'est la Sofina elle-même qui n'ose pas s'opposer à la puissance de la *General Electric* et de ses banques américaines[77]. Dans un domaine où elle se sent très bien préparée sur le plan technique, c'est-à-dire dans celui des téléphones, *Siemens* réussit par

contre à s'implanter solidement dans une partie de l'Argentine et de l'Uruguay[78], après avoir enfin trouvé, à la suite de divergences prolongées, un *modus vivendi* avec l'*ITT* prédominante dans l'ensemble, avec la firme suédoise *LM Ericson* et avec le groupe italo-argentin *Herlitzka*[79]. Il semble donc que ce soit, en fin de compte, le facteur technologique quí, pour l'industrie électrotechnique allemande – en l'absence du soutien financier – redevient primordial, comme dans la phase initiale du développement de l'électricité, jusqu'au début des années 1890.

ANNEXE

*Participation à l'exportation mondiale de produits lectrotechniques (pourcentages)*

| | 1913 | 1925 | 1927 | 1929 | 1931 | 1933 | 1935 | 1937 |
|---|---|---|---|---|---|---|---|---|
| Allemagne | 46,4 | 25,8 | 26,6 | 27,8 | 32,7 | 29,8 | 26,0 | 26,5 |
| États-Unis | 15,7 | 24,9 | 25,2 | 26,5 | 23,2 | 19,8 | 23,5 | 25,9 |
| Grande-Bretagne | 22,0 | 25,1 | 23,2 | 17,6 | 13,8 | 16,2 | 19,5 | 18,9 |
| Pays-Bas | 1,7 | 3,8 | 4,3 | 9,2 | 7,4 | 10,9 | 9,3 | 8,1 |
| France | 4,2 | 5,6 | 3,8 | 3,3 | 4,5 | 4,7 | ? | 2,3 |
| Suède | 2,0 | 2,8 | 3,9 | 3,1 | 3,5 | 2,9 | ? | 2,9 |
| Suisse | 3,5 | 3,4 | 3,2 | 3,1 | 3,5 | 4,0 | ? | 2,4 |

*Source:* A. R. GLARDON, *Die deutsche Elektroindustrie und der Absatz ihrer Erzeugnisse in der Nachkriegszeit*, Hamburg, 1933, p. 17; P. CZADA, *Die Berliner Elektroindustrie in der Weimarer Zeit*, Berlin, 1969, p. 317; K. SCHRÖTER, *Die aussenwirtschaftliche Stellung der deutschen Elektroindutrie*, thèse dactylographiée, université de Francfort, 1940, p. 52 s.

## Notes

[1] Cf. V. Schröter, *Die deutsche Industrie auf dem Weltmarkt 1929 bis 1933*, Frankfurt a.M./Bern/New York/Nancy, 1983, p. 82 s. (une étude excellente et fondamentale pour toute la question).

[2] British Electrical & Allied Manufacturers' Association. ERconomic and Statistical Department (ed.), *Combines and trusts in the electrical industry. The position in Europe in 1927*, London, 1927, p. 97.

[3] Siemens-Museum, München, Firmenarchiv (par la suite abrégé en ASM: SAA 11/Lb 581 (Nachlass Liedtke), relevé du 4.6.1913.

[4] Calculé d'après A. R. Glardon, *Die deutsche Elektroindustrie und der Absatz ihrer Erzeugnisse in der Nachkriegszeit*, Hamburg, 19-3, p. 15; F. Gapinski, *Die Stellung der deutschen Elektro-Industrie innerhalb der international Elektro-Wirtschaft in der Gegenwert*, thèse de doctorat, université de Cologne, Berlin, 1931, p. 85, 90.

[5] E. Brandstetter, *Finanzierungsmethoden in der deutschen elektrotechnischen Industrie*, thèse de doctorat, université de Giessen, 1930, p. 4 (ce chiffre comprend probablement aussi la société Lahmeyer).

[6] F. Pinner, *Emil Rathenau und das elektrische Zeitalter*, p. 314 (le chiffre pour Siemens comprend probablement aussi la compagnie Bergmann).

[7] P. Czada, *Die Berliner Elektroindustrie in der Weimarer Zeit*, Berlin, 1969, p. 279.
[8] A. Gebhardt, *Die Expansion der amerikanischen Elektro-Konzerne in Europa*, thèse de doctorat, université de Heidelberg, Wertheim/Main, 1932.
[9] Cf. en particulier G. Siemens, *Der Weg der Elektrotechnik. Geschichte des Hauses Siemens*, vol. 1, Freiburg i.B./München, 1961; S. Von Weiher/H. Goetzeler, *Weg und Wirken der Siemens-Werke im Fortschritt der Elektrotechnik, 1847-1972*, München, 1972-1973; J. Kocka, *Unternehmensverwaltung und Angestelltenschaft am Beispiel Siemens, 1847-1914*, Stuttgart, 1969.
[10] Allgemeine Elektrizitätsgesellschaft (ed.), *50 Jahre AEG*, manuscrit imprimé, Berlin, 1956; G. Siemens, *op. cit.*, vol. 1, p. 108 ss; F. Pinner, *op. cit.*, p. 80 ss. A Broder, «L'expansion internationale de l'industrie allemande dans le dernier tiers du XIXe siècle: le cas de l'industrie électrique: 1880-1913», *Relations internationales*, n° 29, printempt 1982, p. 65-87, surtout p. 70 s.
[11] Cf. J. Kocka, «Siemens und der aufhaltsame Aufstieg der AEG», *Tradition*, a. 17 (1972), p. 125-142; pour le concept de «managerial enterprise» cf. A. D. Chandler, *The United States: Seedbed of managerial capitalism*, dans A. D. Chandler/H. Daems (eds), *Managerial hierarchies. Comparative perspectives on the rise of the modern industrial enterprise*; Cambridge, Mass./Londres, 1980, p. 9-40.
[12] Pour la crise de 1901-1902 dans l'industrie électrotechnique, cf. J. Loewe, «Die elektrotechnische Industrie», dans *Die Störungen im deutschen Wirtschaftsleben während der Jahre 1900 ff.* (Schriften des Vereins für Socialpolitik, vol. 107, t. 3), Leipzig, 1903 p. 77-155.
[13] Cf. G. Eibert, *Unternehmenspolitik Nürnberger Maschinenbauer (1835-1914)*, Stuttgart, 1979, p. 256 ss, 372.
[14] G. Siemens, *op. cit.*, vol. 1, p. 228 ss; J. Kocka, *Unternehmensverwaltung, op. cit.*, p. 324 ss.
[15] *50 Jahre AEG, op. cit.*, p. 155 s.
[16] *Ibidem*, p. 179 s.
[17] J. Kocka, *Unternehmensverwaltung, op. cit.*, p. 333 s.; G. Siemens, *op. cit.*, vol. 2, p. 9 s.
[18] J. Kocka, *Unternehmensverwaltung, op. cit.*, p. 570, 563 (chiffres arrondis).
[19] K. Wilhelm, *Die AEG*, Berlin, 1931, p. 69.
[20] Cf. P. Hertner, «Fallstudien zu deutschen multinationalen Unternehmen vor dem Ersten Weltkrieg», dans N. Horn/J. Kocka, *Recht und Entwicklung der Grossunternhemen im 19. und frühen 20. Jahrhundert*, Göttingen, 1979, p. 388-419, en particulier p. 410 ss.
[21] E. Waller, *Studien zur Finanzgeschichte des Hauses Siemens* (manuscrit dactylographié conservé dans SMM), t. IV, p. 207.
[22] J. Kocka, *Unternehmensverwaltung, op. cit.*, p. 563; S. Von Weiher/H. Goetzeler, *op. cit.*, p. 62.
[23] G. S. Holzer, *The German electrical industry in Russia: From economic entrepreneurship to political activism, 1890-1918*, thèse de doctorat, University of Nebraska, 1970, p. 23; S. Von Weiher, *Die Entwicklung der englischen Siemens-Werke und des Siemens-Überseegeschäfts in der zweiten Hälfte des 19. Jahrhunderts*, thèse de doctorat (dactylographiée), université de Fribourg-en-Brisgau, 1959, p. 44 s.
[24] *50 Jahre AEG, op. cit.*, p. 158 s.; H. Von Sothen, *Die Wirtschaftspolitik der Allgemeinen Elektrizitäts-Gezellschaft*, thèse de doctorat, université de Fribourg-en-Brisgau, 1915, p. 36.
[25] Zentrales Staatsarchiv Potsdam, Auswärtiges Amt Nr. 5804, Bl. 120 a.
[26] *50 Jahre AEG, op. cit.*, p. 184.
[27] V. Djakin, «Zur Stellung des deutschen Kapitals in der Elektroindustrie Russlands», *Jahrbuch für Geschichte der UdSSr und der volksdemokratischen Länder Europas*, 9 (1966), p. 121-143, et en particulier p. 142 s. P. Lanino, *La nuova Italia industriale*, vol. 2, Roma, 1916, p. XXVIII; G. Jacob-Wendler, *Deutsche Elektroindustrie in Lateinamerika. Siemens und AEG (1890-1915)*, Stuttgart, 1982, p. 69.

[28] Pour le cas de l'industrie électrotechnique britannique, cf. I. .C. R. Byatt, *The British electrical industry, 1875-1914*, Oxford, 1979, p. 166 ss.

[29] Cf. H. Von Sothen, *op. cit.*, p. 17 ss.; C. Matschoss, «Die geschichtliche Entwicklung der Allgemeinen Elektricitäts-Gesellschaft in den ersten 25 Jahren ihres Bestehens», *Beiträge zur Geschichte der Technik und Industrie*, vol. 1 (1909), p. 53-72, surtout p. 67 s.; pour Siemens, cf. J. Kocka, *Unternehmensverwaltung, op. cit.*, p. 372 ss.

[30] Pour le paradigme de l'internationalisation, cf. R. H. Coase, «The nature of the firm», Economica, a. 1937, p. 386-405; O. E. Williamson, «The modern corporation: Origins, evolution attributes», *Journal of Economic Literature*, a. 1981, p. 1537-1568.

[31] J. H. Dunning, *International production and the multinational enterprise*, London, 1981, p. 25 ss.

[32] Zum 70. Geburtstag Dr. Felix Deutsch, *Die AEG-Umschau*, 16 mai 1928, p. 33 s.

[33] Cf. par exemple V. Schröter, *op. cit.*, p. 449 s.; l'argument «défensif», pourtant dans le contexte d'une stratégie multinationale, a été utilisé d'abord par R. Vernon, «International Investment and international trade in the product cycle», *Quarterly Journal of Economics*, 80 (1966), p. 190-207.

[34] Cf. R. Liefmann, *Beteiligungs- und Finanzirungsgesellschaften. Eine Studie über den modernen Kapitalismus und das Effektenwesen*, Jena, 1913, p. 103 s.; F. Fasolt, *Die sieben grössten deutschen Elektrizitäsgesellschaften*, thèse de doctorat, université de Heidelberg, Borna bei Leipzig, 1904, p. 31 s.; M. Jörgens, *Finanzielle Trustgesellschaften*, thèse de doctorat, université de Munich, Stuttgart, 1902, p. 117 ss.

[35] Cf. K. Hafner, *Die schweizerischen Finanzierungsgesellschaften für elektrische Unternehmungen*, thèse de doctorat, université de Fribourg (Suisse), Genève, 1912, p. 30 ss; M. Jörgens, *op. cit.*, p. 73 ss; R. Liefmann, *op. cit.*, p. 504 ss.

[36] Cf. P. Hertner, *Il capitale tedesco in Italia dall'Unità alla Prima Guerra Mondiale*, Bologne, 1984, p. 214 ss.

[37] Cf. en particulier A. Strobel, «Die Gründung des Züricher Elektrotrusts. Ein Beitrag zum Unternehmergeschäft der deutschen Elektroindustrie», dans H. Hassinger (ed.), *Geschichte-Wirtschaft-Gesellschaft. Festschrift für Clemens Bauer zum 75. Geburtstag*, Berlin, 1974, p. 303-332; H. Grossmann, *Die Finanzierungen der Bank für elektrische Unternehmungen in Zürich*, thèse de doctorat, université de Zurich, 1981; K. Hafner, *op. cit.*, p. 32 ss; R. Liefmann, *op. cit.*, p. 463 ss.

[38] Cf. E. Waller, *op. cit.*, t. IV, p. 19 s.

[39] P. Hertner, *Il capitale tedesco, op. cit.*, p. 223.

[40] Cf. K. Hafner, *op. cit.*, p. 57 ss.

[41] Cf. R. Liefmann, *op. cit.*, p. 505 s.; 456 ss; M. Dumoulin, *Italie-Belgique: 1861-1915. Relations diplomatiques, culturelles et économiques*, thèse de doctorat, université catholique de Louvain, 3e partie, vol. 2, Louvain-la-Neuve, 1981, p. 837 ss; A. Confalonieri, *Banca e industria in Italia, 1894-1906*, vol. 3, Milano, 1976, p. 259 ss.

[42] K. Singer, «Die deutsche Auslandsverschuldung», *Wirtschaftsdienst*, n° 27 du 7 juillet 1922, p. 660-661.

[43] Cf. l'excellent résumé de la question par V. Schröter, *op. cit.*, p. 114 ss, et en particulier p. 118.

[44] *Ibidem*, p. 220, et surtout «Deutschland in der internationalen Kapitalverpflechtung», *Wirtschaft und Statistik*, 10 (1930), p. 890-895.

[45] L. Michel, *L'électricité en Belgique. Organisation économique et financière* (université de Liège. École supérieure des sciences commerciales et économiques), Liège s.d «1936?», p. 182.

[46] B. Fehr, *Zusammenschluss und Finanzierung in der Elektrizitätindustrie*, thèse de doctorat, université de Berne, Berne, 1939, p. 46 s.; *Bericht des Verwaltungsrates der Bank für elektrische Unternehmungen an die ordentliche Generalversammlung der Aktionäre vom*

*31.Oktober 1921*, p. 4 ss; *Der AEG-Konzern 1928* (Das Spezial-Archiv der deutschen Wirtschaft, Sonderausgabe), Berlin, 1928, p. 18; *50 Jahre AEG, op. cit.*, p. 235.

[47] Cf. E. Waller, *op. cit.*, t. V, p. 192 ss; *50 Jahre AEG*, p. 233 s.

[48] G. Siemens, *op. cit.*, vol. 2, p. 268.

[49] H. Schröter, «Siemens and Central and South-East Europe between the two world wars», dans A. Teichova et P. C. Cottrel (eds.), *International business and Central Europe, 1918-1939*; Leicester/New-York, 1983, p. 173-192, et en particulier p. 18 s.; V. Schröter, *op. cit.*, p. 437 ss.

[50] V. Schröter, *op. cit.*, p. 441.

[51] *Ibidem*, p. 415, note 247; H. Schröter, *op. cit.*, p. 180.

[52] *Der AEG-Konzern 1928, op. cit.*, p. 19.

[53] E. Waller, *op. cit.*, t. V, p. 2313; V. Schröter, *op. cit.*, p. 424.

[54] V. Schröter, *op. cit.*, p. 424.

[55] *Ibidem*, p. 443 s.

[56] *Ibidem*, p. 420 ss; E. Waller, *op. cit.*, t. V, p. 56 s.; J. D. Scott, *Siemens Brothers, 1858-1958*, London, 1958, p. 230.

[57] K. Pohl, *Weimars Wirtschaft und die Aussenpolitik der Republik, 1924-1926*, Düsseldorf, 1979, p. 127 ss; W. Link, «Der Amerikanische Einfluss auf die Weimarer Republik in der Dawesplanphase (Elemente eines 'penetrierten Systems'», dans H. Mommsen/D. Petzina/B. Weisbrod (eds.), *Industrielles System und politische Entwicklung in der Weimarer Republik*, réimpression Düsseldorf, 1977, p. 485-498, surtout p. 488 s.

[58] V. Schröter, *op. cit.*, p. 258; E. Hess, *Elektropolitik und Weltvertrustung*, Amsterdam, 1931, p. 9.

[59] E. Hess, *op. cit.*, p. 8.

[60] E. Waller, *op. cit.*, t. V, p. 151 ss; V. Schröter, *op. cit.*, p. 394 ss; pour la banque Dillon Read, cf. V. P. carosso, *Investment banking in America. A history*, Cambridge, Mass./London, 1970, p. 289 ss.

[61] K. Wilhelm, *op. cit.*, p. 90; V. Schröter, *op. cit.*, p. 258.

[62] E. Waller, *op. cit.*, t. V, p. 80 s.; V. Schröter, *op. cit.*, p. 258 s.

[63] V. Schröter, *op. cit.*, p. 263.

[64] Ibidem, p. 332 ss; M. Wilkins, *The maturing of multinational enterprise: American business abroad from 1914 to 1970*, Cambridge, Mass./London, 1974, p. 70 s.; 130 s.

[65] E. Waller, *op. cit.*, t. V, p. 189.

[66] G. Siemens, *op. cit.*, t. 2, p. 260 ss; K. Wilhelm, *op. cit.*, p. 94.

[67] E. Waller, *op. cit.*, t. V p. 95, V. Schröter, *op. cit.*, p. 150 s.

[68] V. Schröter, *op. cit.*, p. 150 s.; E. Hess, *op. cit.*, p. 25 ss; *Der AEG-Konzern 1928, op. cit.*, *passim.*

[69] Cf. P. Hertner, «How they changed their strategy: German electrotechnical industry in the Italian market before 1914 and between the two wars», Contribution to the conference on *The early phase of multinationale enterprise in Germany, France and italy*, Florence, 17-19 October, 1984, European University Institute, Florence «EUI Colloquium papers 159/84 (col. 28».

[70] ASM, SAA 11/la 137 (Jessen), Bd. 11; A. R. Glardon, *Die deutsche Elektroindustrie und der Absatz ihrer Erzeugnisse in der nachriegszeit*, Hamburg, 1933, p. 61, 117.

[71] Cf. surtout le cas des *Officine Lombarde Apparecchi di Precisione* (OLAP), fondées en 1927 (ASM, SAA 68/li, 182: compterendu sur la SAM du 12 décembre 1945; *ibidem*, SAA 47/LK 843: compterendu sur Siemens en Italie du 9 septembre 1944).

[72] ASM, SAA 68/Li 144: Spanien («Werk Conellá»); pour le tarif douanier de 1906, cf. J. Harrison, *An economic history of modern Spain*, Manchester, 1978, p. 84 s.

[73] ASM, SAA 54/Ld 830; «AEG und SSW in Spanien 1921-1931»; pour une éventuelle fusion entre *AEG* et *Siemens* en Allemagnemême, cf. V. Schröter, *op. cit.*, p. 143, et ASM,

SAA 4/Lt 398 (II 10): «Niederschrift über eine Erörterung geschäftspolitischer Fragen» du 5 janvier 1929.

[74] E. Waller, *op. cit.*, t. V, p. 226.

[75] Cf. G. Jacob-Wendler, *Deutsche Elektroindustrie in Lateinamerika. Siemens und AEG (1890-1914)*, Stuttgart, 1982, p. 71 ss; *Elektrotechnische Zeitschrift*, a. 1912, p. 283; *ibidem*, a. 1913, p. 786, 856.

[76] Cf. E. Hess, *op. cit.*, p. 21 s.; B. Fehr, *op. cit.*, p. 48 s.

[77] ASM, SAA 20/La 969-971; M. Wilkins, *The maturing of multinational enterprise: American business abroad from 1914 to 1970*, Cambridge, Mass./London, 1974, p. 131 s.

[78] ASM, SAA 4/Lf 701, SAA 68/Li 260: «Argentinien 1897-1970»; *ibidem*: «Lateinamerika. Länder R-Z» «Uruguay»; cf. V. Schröter, *op. cit.*, p. 336 ss.

[79] ASM, SAA 11/Lg 701 (Franke); SAA 4/Lf 657; SAA 11/La 137 (Jessen), Bd. 5; cf. A. Attman/J. Kuuse/U. Olsson, *LM Ericsson 100 years*, vol. 1, Örebro, «1976», p. 274 s., vol. 2, «1976», p. 25.

# The Beyen Plan and the European Political Community

RICHARD T. GRIFFITHS AND ALAN S. MILWARD
Amsterdam and London/Florence

The Netherlands had agreed to participate fully in the discussions to create a European army only in October 1951, only with great reluctance, only by a decision of a divided government and only under strong American, French and German pressure. When agreement to the European Defence Community seemed the only way to ensure an effective land defence of the Netherlands, and only then, did the Dutch accept the management of a common defence as a reason for common European political institutions. The Dutch interest in such institutions was economic. More precisely it was commercial. Foreign commodity trade accounted for more than a third of national income. Most of it was with western Europe. Foreign invisible earnings had historically always weighed heavily in the balance of payments and there was every reason to suppose they would do so in the future. The search for higher levels of income and prosperity demanded maximum access to European markets and a better regulation of international payments and capital movements. This was the reason for seeking common European action. About the framework within which that common action should be taken Dutch governments were in a much greater state of confusion. The Stikker proposals of June 1950 had envisaged common action to reduce both tariffs and quantitative trade restrictions in the wide framework of OEEC and had been prepared for this purpose to concede a marginally greater degree of executive and managerial power to the Executive Committee and Council of OEEC, particularly in order to provide 'compensations' from a common monetary fund to governments whose industries were particularly damaged by this process. After December, however, faced with strong opposition from the Italian government and the likelihood that the French government was sheltering behind the Italian position, the Dutch had more or less abandoned this OEEC framework in the hope of getting better results from GATT. This was even less successful. The disturbances of the Korean War meanwhile had led to the suspension, nominally temporary, of many of the trade liberali-

sation measures which ECA had squeezed out of OEEC in return for Marshall Aid.

Any plan in GATT which required the agreement of all members to tariff reduction, as had the Pflimlin proposals after the Torquay conference in 1951, was more or less doomed to failure because of the presence there of underdeveloped economies also. Europe looked the only feasible framework for Dutch action. But that meant encountering French intentions to maintain protection either by tariffs or quantitative restrictions. The extension of preferences by the existing processes of tariff bargaining so as to create a preferential tariff zone in western Europe, along the lines of the Italian proposals in summer 1950, would only be economically helpful to the Netherlands if it produced results very quickly, in the first round of negotiations in fact. It would not do anything about quota restrictions. It might well not get the agreement of GATT. A gradual resumption of the process of trade liberalisation, that is to say joint removal of quotas under OEEC supervision, would presumably again encounter the same resistance. It would also not serve the purpose of providing sure and growing markets for Dutch agricultural exports, because there was no likelihood that countries would be any more willing to reduce their barriers against agricultural imports than they had been in 1950. There were therefore certain persuasive arguments that nothing could be achieved without the creation of some new European body with a political mechanism which might force a process of change in the desired direction.

The High Authority of the ECSC was of course just such a body. Its importance for the Dutch economy was very limited. Had it been greater the Netherlands would probably never have accepted such a form of political machinery. The small coal sector and the even smaller steel sector in the Dutch economy both worked at relatively high levels of efficiency and were in no serious economic or political danger of having to make much adjustment as a result of the institution either of the common market or the High Authority. Even so the political shape of the institutions of the Community had been greatly changed in the negotiations because of Dutch pressure. The Assembly, the Council of Ministers, the advisory committee, the Court of Justice, all had been strengthened in their position relative to the High Authority, because of Dutch opposition to the apparently uncontrolled technocratic, interventionist form which the French proposals had originally wished to give to the High Authority. The Netherlands had sternly resisted all attempts to transpose the political concepts and machinery of the Monnet Plan to a European level, and there was great opposition inside the Dutch government to repeating this experiment, even in the modified form in which it had come into being, for any other sector of the economy.

This had become clear in the cabinet struggles over Mansholt's

proposals for a similar high authority to regulate western Europe's intra-trade in agricultural products. Like the ECSC such a solution had two glaring weaknesses from a Dutch point of view, leaving aside entirely the question of how acceptable it was to other nations. Firstly it placed the Netherlands in a separate framework for political action from the United Kingdom, when the United Kingdom was one of its two most important commercial partners and a crucial military ally. Secondly it placed the Netherlands in a political framework where French influence would be preponderant. And since French policy was to strengthen the domestic methods of economic management by extending them through western Europe this augured a future of trade controls, production agreements and market sharing agreements which, except in agricultural products, was not exactly what the Netherlands was looking for and presented obvious dangers of subordinating Dutch domestic economic policies to the French interest. These French pressures were obvious in the negotiations over the EDC, especially in the French insistence on coordinating western European armaments production by a centralised Commissariat directed by Monnet's successor in the Commissariat au Plan. It was this which had brought the EDC negotiations to deadlock at the end of 1951. Furthermore France showed little interest in any European association wider than the six members of the ECSC. It was a cardinal point of French policy that if new European organizations with real powers came into existence their purpose must also be that of retaining the West German republic in the integrated western European framework. The 'six' had been painfully constructed for this purpose in the iron and steel sector only after its construction had failed in the foreign trade and payments sector. In effect the rationale for now once more confining to the six members of the ECSC the development of foreign trade rules in western Europe was a purely pragmatic one. Its damages were mitigated by the existence of the European Payments Union which continued to operate its own trade rules in the wider OEEC framework. In that sense demanding a liberal common market as part of the federated western Europe which the EDC was supposed to bring about could be construed as exercising pressure to make the OEEC trade rules more liberal. It served the purpose of strengthening the Franco-German peace treaty and association embodied in the Schuman Plan and the Treaty of Paris as well. This was certainly of overwhelming importance to the Netherlands, but the problem of Dutch foreign trade was one which went far beyond the confines of the six and was thus conceived as being of even more fundamental importance.

It was also true that the bargaining position of the Netherlands was stronger in the six than in a wider setting, because, as the Schuman Plan negotiations had proved, French policy was unrealisable without some agreement with the Benelux countries to act in concert. In fact in the wider forum of the OEEC the Netherlands had had little influence on the

process of trade liberalisation and the tortuous negotiations on agricultural trade in the OEEC Food and Agriculture Committee had produced a state of total disillusionment. A reasonable weighing up of the Dutch chances of improving the international framework of the domestic Dutch economy after 1950 would have pointed to the six as a more promising framework for political action, although the chances of success could not be viewed optimistically. Furthermore, if the ECSC was actually going to be extended into a European Political Community (EPC) in order to bring the European army under some form of political control, everything spoke in favour of the Netherlands at least attempting to gain some concessions in return for its adhesion to this new European political body and even trying to convert it into a body which would have an economic, rather than a merely political, purpose beyond that inherent in the ECSC.

When the foreign ministers finally produced the draft treaty for the EDC in May 1952 the second paragraph of Article eight stressed that the proposed constitution of the EDC was only temporary and that it should eventually be replaced by a new European organisation. This organisation would clearly have to be related in some way to the ECSC either by assimilation into it or by a reconstruction of the first Community. The existing Court of Justice was to serve also for the EDC. On the other hand the members of the proposed parliamentary assembly would be the members of the Council of Europe with some additions. Everything pointed towards a weakening of the supranational powers of the ECSC. Benelux pressure had ensured that the European Commissioners for Defence would in no sense have the same degree of independence from national governments as the High Authority. The possibility therefore existed that because the new European organisation which was to be created would not have the same independence or strongly interventionist character as the High Authority, the other nations might not object if it were enjoined with carrying out a programme of removal of barriers to trade as well. The investigation of this new form of European organisation should, the draft treaty insisted, be undertaken within six months of the ratification of the treaty by the proposed EDC assembly. The Italian government had then proposed for the meeting of the foreign ministers in September in Luxembourg that the parliamentary assembly be established immediately in *ad hoc* form for one specific purpose, to study the possible formation of a European Political Community at once without waiting for the ratification of the EDC treaty. Any modifications to the purely political nature of the proposed EDC would therefore have to be proposed in September at the latest.

The proposals had to be formulated by a new government and a new foreign minister. Stikker's efforts had all been in the context of the OEEC and his dislike of any arrangement, political or economic, which disassociated the Netherlands from Britain had been strong. After the general

election he was replaced by two foreign ministers, Jan Willem Beyen and Joseph Luns. Beyen was specifically charged with multilateral relations, which included international commercial relations, but Luns was in charge of Benelux affairs. As minister of economic affairs Zijlstra had replaced van der Brink. Drees remained as prime minister and Mansholt stayed as minister of agriculture. The initiative to turn the proposed EPC into an economic organisation as well came from Beyen and Luns appears to have had very little to do with it. The change of policy from Stikker to Beyen was less a matter of personality than of reaction to events. For both, the central issue was the reduction of trade barriers. Stikker having made so little progress in OEEC it was normal enough to go about it in a different direction whoever was in charge. It is clear, however, from subsequent events after the failure of the EDC treaty, that Beyen was prepared to push harder and go further than his civil servants towards some form of 'integration' with the six, always providing that it served Dutch commercial purposes which he defined no differently from Stikker. He was simply more optimistic about the chances than were his advisers and most other members of the government and less deterred by the possible dangers. Although Dutch policy followed a coherently rational line throughout the changes in government and was essentially a response to the events outside the Netherlands the difference that an energetic minister with a will to push his own line could make was certainly observable, as it was also to be in 1955 after the complete collapse of the proposed EPC.

That being said, much more important than personality was the repercussive nature of policy formulation in the Netherlands, as in all the member states. The French proposals in 1950 for a closer American integration into NATO planning had produced the American proposals for German rearmament in a European Defence Community. The French and German attempts to construct a purely political body to oversee a common defence forced the Dutch to the position that if they were forced to accept a European Political Community it should have an economic purpose. No side actually wanted a further European organisation of any kind, much less one with real powers, for its own sake. But if the Defence Community, which by 1952 only the Germans and Americans really seemed to want anyway, was the only way forward towards solving the political problem of combining German rearmament, French security and America's foreign policy objectives in Europe, it was now not going to be achieved unless the Political Community, which in turn was the price paid for ensuring the control of the Federal Republic within the Defence Community, also helped to solve the international economic problems of the Netherlands. The Dutch government would have preferred to have solved these in a wider framework and without a further level of European organisation had the possibility existed. Adding to the superstructure of an unwanted house built on no foundations of real political will or purpose was only likely, of

course, to make its collapse more certain. But what else can small states do but accept the repercussive nature of foreign policy and hope that their interests will not be entirely dismissed? It is hardly up to them to change the agenda for international negotiation. In this way the structure of the European Political Community was to become even more gothic, a soaring mass of complicated and ill-connected aspirations to the solution of problems which had only one thing in common, the fact that they could not be solved within the national frontiers, but were otherwise absolutely different in nature.

This was well expressed by the Dutch Secretary-of-State for Foreign Affairs, Hendrik Boon, who considered that the Italian resolution seemed to be working from the premise that the establishment of a 'political authority' would be the panacea for all the troubles facing Europe but he wondered whether the reverse might not be true, – that in working towards and propagating a political community the drive towards solving fundamental economic issues might be weakened even further. Besides that, he felt that the Dutch should resist any change in the EDC treaty which might imply the eventual devolution of any power to a supranational organisation without arrangements to protect the individual states' interests. Finally he argued that if it was intended to abandon the functional approach to integration, this should only take place after all the other options had been properly studied by national governments[1]. In August a state commission was set up to advise the government on the whole question of European integration. Beyen, however, was determined to stake some sort of claim at Luxembourg when the foreign ministers met.

The claim was vague, as it could only be in the circumstances. The *Assemblé ad hoc*, he argued, should consider as well as the political form of the EPC the question of economic integration. Van Zeeland supported this position completely and de Gasperi had no real objections. This study, Beyen also argued, should be implemented under the guidance of the foreign ministers who would provide guidelines. Again this idea was supported so that Beyen had at least altered the situation in the sense that the foreign ministers could lay down, if they were capable of agreement, economic guidelines for the future activity of a Political Community[2].

In October the State Commission which had been established to advise the government on the question of European integration produced the results of its deliberations on the Luxembourg Resolution. It considered that in the area of economic integration (and especially in agriculture) it would be difficult for the Netherlands to obtain quick results. Indeed it was likely that the negotiations would get little further than an agreement that the EPC would concern itself with economic questions. It might possibly even get round to deciding which economic questions to discuss. But the whole process could be time-consuming. Thus, the Dutch should not enter the EPC unless satisfactory progress were made on the

economic front. What was needed was for an economic study group to be set up with a detailed list of instructions, since only in this way could it get on with serious activity in the direction which the Dutch wanted[3].

In November, Beyen formulated what he believed should be the basis for the Dutch standpoint on European integration. He painted a picture of Europe threatened externally by military pressure and internally by communism and fascism. To cope with these pressures an increase in production and productivity was necessary, but this was impossible as long as Europe was splintered into small markets by trade discrimination and monetary uncertainty. So far the external threat had made clear that military cooperation was necessary and the realisation had grown that this required a measure of political cooperation. However the question should be asked whether political integration without economic integration made any sense. 'Political integration which has no content other than making possible coordinated military action and organising the production and marketing of certain important raw materials can only bring about a very limited unity'[4]. To a certain extent these remarks must have been directed against his cabinet colleagues. It was Drees who had denied the seriousness of the military threat to western Europe in 1950. The proposals which Beyen went on to make were even less likely to meet with universal accord. Because it would be a waste of energy to attack all barriers to trade in whatever form the eventual goal must be a customs union. Only this could guarantee the institutionalisation of freedom to trade across frontiers. The customs union would have to be attained in stages as and when improvements in the separate national economic situations allowed. If it were implemented at once economic disturbances would produce a resurgence of protection. Here, presumably, he had in mind the sad fate of the OEEC trade liberalisation programmes. A programme of this kind implied a supranational authority to direct it. The first steps should be taken in those sectors where output was most restricted by trade barriers. The cabinet should set up a special commission under his chairmanship to work out these proposals in greater detail.

The timetable foresaw a further meeting of the foreign ministers at Rome in February 1953. By this time the Beyen proposals had taken a final and slightly altered shape. The form of the proposed customs union had become more specific. No longer was Beyen content to leave the implementation of the customs union to any supranational power. Both the target date and the timetable for achieving it were to be specified in the treaty. Behind the common external tariff there would also have to be a programme for removing quantitative trade restrictions, transport monopolies and discriminations and barriers to invisible transactions. The Community should administer a series of safety clauses under fixed values. To help in this the familiar idea from 1950 of a 'European Fund' was resurrected, to be deployed by the Community in cases where there were

'fundamental difficulties'. The relationships with non-members of the Community should be regulated from the start in the terms of the treaty[5]. It is hard to believe that after the start of 1953 there was any real chance of the EDC treaty being passed in the French parliament. The decision of the Mayer government to seek additional protocols to the treaty had begun to alter its nature without making it any more likely that it would finally be acceptable to the national assembly. The new Dutch proposals only exacerbated the position, putting ratification and a political settlement of the Franco-German question further at risk. Beyen's reaction of course was that it was doubtful if a parliamentary majority for ratification could be found in the Netherlands without including in the treaty a proposition for a common market. Belgium and Italy both supported this position and the conference communiqué eventually contained a section reaffirming the Luxembourg resolutions and agreeing to study the possibility of implementing a common market within the terms of that resolution[6].

The draft treaty was handed over by the *Assemblé ad hoc* to the foreign ministers in Strasbourg in March at their next meeting for their further consideration and in May they met in Paris to pronounce on it.

The treaty envisaged a directly elected parliament, elected at first on the basis of the separate national systems but after five years by a common European electoral system to be established by the parliament itself. It would have 15 seats specially reserved for the Saarland, three more than for Luxembourg. Above it would be a Senate, whose members would be directly appointed by national parliaments. The Senate would appoint some of the members of a European Executive Council and its chairman would appoint the others. It could be removed by a motion of no-confidence from either chamber of the Parliament. There should also be a Social-Economic Council comprising various interest groups which could present advice to the Executive but which also had rights of initiative. A Council of National Ministers would have the task of harmonising relations between the Executive and national governments and exercising any further powers delegated to it by the Senate. Finally there would be a Court with responsibilities for ruling on interpretations of the Treaty and on questions of constitutional abuse.

Exactly what this impressive constitutional construction was supposed to do was rather less clear. On foreign policy the only rights attributed to the parliament were the approval of treaties with third countries and association agreements. The Council of Ministers was to coordinate a common policy in international conferences and to be receptive to proposals and advice from the Executive and parliament. The Community was also to have its own budget, but its size and financing was left ultimately to the Executive and had to be approved by a unanimous decision in the Council of Ministers. The amendments which parliament could make had to be within the framework of the total expenditure agreed. Again, the

Community had the task of progressively creating a common market but on the basis of measures proposed by the Executive, approved by a unanimous vote in the Council of Ministers and by a majority of both Chambers of Parliament. There was also to be an adaptation fund under parliamentary control. Finally, after two years, the EPC would completely take over the control of the ECSC and the EDC[7]. The economic committee of the Council of Europe reviewing its work in May frankly conceded 'The European Community to be set up is endowed in the Draft Treaty with economic powers which would make possible a progressive development towards a common market. The Draft Treaty does not, however, contain any obligatory provisions concerning economic integration, and the safeguards foreseen are such that if they are exploited they would make impossible any substantial progress towards achieving the objective proclaimed. One might almost say that the economic provisions contained in the Draft Treaty presented the only possible compromise that could be reached by a majority of the *Assemblé ad hoc*, rather than a unanimous or a whole-hearted support for immediate action towards the establishment of a common market (by) the six countries concerned'[8].

Meanwhile at the end of April, the Dutch Commission set up by Beyen had formulated its first conclusions. The economic sub-committee, having pointed out that the realisation of a common market was not just a question of removing trade barriers but also of harmonising monetary and social policies which artificially distorted national structures, went on to criticise the Beyen proposals as being inadequate by themselves for the realisation of their own ultimate goals. 'The opening of frontiers', it warned, 'is almost impossible as long as a certain measure of coordination has not been achieved in the areas of general economic, financial, monetary and social policy'. It recommended that article 82 of the draft treaty be amended accordingly. It also wanted to see a synchronic treatment of tariffs, quotas and a common external trade regime. But it argued against fixing a definitive end date partly because the realisation would depend on the progress towards harmonisation elsewhere, partly because no one was likely to agree to it and partly because it was outside the Dutch principles – 'it is about integration and not a unitary state'. Apart from that it recommended some procedural changes made in the way the safeguard clauses were dealt with and the way the fund was administered[9]. The institutional sub-committee was much less iconoclastic as far as the main lines of Dutch policy were concerned. It warned against the early subsumation of the ECSC and EDC into a new political community, because it would require a substantial modification of both treaties and also because it would deflect attention from the wider economic task. On the other hand it recognised the force of those arguments which said that if one were striving for political economic integration it was illogical to maintain the existence of separate functional communities. Perhaps, it suggested, a

compromise could be found along the lines of a partial subsumation or by aiming for a later date. Secondly, the subcommittee was opposed to the senate electing the chairman of the Executive and the chairman in turn appointing the rest. It preferred to leave both questions to the Council of Ministers. It was not too enamoured either with the balance within the senate, where it preferred an equal representation for all states, but it felt that to achieve this, the Dutch might have to concede a form of representation in the parliament based on the size of population, which would leave small states in a worse position than envisaged by the treaty. It made no recommendation of its own on the question of direct elections[10].

Beyen's simple formulation of economic integration virtually exclusively in terms of a customs union was challenged by the higher civil servants represented on the advisory committee as not going far enough. More serious, however, was the challenge from within the cabinet that the Dutch had gone far enough already. This group was led by the Prime Minister, Willem Drees. He stated quite bluntly that the Six were too unbalanced to form a political community. Political instability hamstrung the decision-making process and there was, besides, a lack of political will for its creation. There was still absolutely nothing to show for the Dutch initiatives to date. And finally, he argued, if the Six did manage to create the whole institutional structure, it would probably fail anyway. To this Beyen replied equally frankly that if the government had felt in its heart that nothing would come of the negotiations it should have said so before they began. In this he was supported by Mansholt who argued that the Netherlands should not pull back from the course upon which it had embarked. Finally Drees backed down but at the same time made it absolutely clear that the creation of a customs union was the minimum and absolute condition for Dutch participation in the EPC. If all that resulted from the negotiations was an involved institutional structure with no advantage for the Netherlands, he reserved the right to resign from the cabinet[11].

In formulating the Dutch reply to the *Projet de Traité*, Beyen therefore ignored the advice of the economic sub-committee and insisted that a date be stipulated in the treaty by which all tariff and quantitative restrictions on intra-trade should be abolished and a common external trade regime should be established. He did, however, take up the other points. In view of the disinterest in cabinet in the political motives for the EPC, although the Dutch memorandum could have dealt with *any* point in the treaty, it made no mention of the institutional clauses at all[12]. Beyen explained later to cabinet that the memorandum had been designed to keep the initiative in Dutch hands, any institutional points he could respond to verbally. Moreover, although he did not anticipate a detailed discussion of the treaty at the meeting of the six in Paris, he explained that he had drafted a number of economic articles for an eventual treaty which he would table *if*

*necessary*. He was advised only to proceed with such a text with extreme caution. The other main conclusion of cabinet was that the draft treaty should not be referred back to the *Assemblé ad hoc*, as the Germans wanted, but should be dealt with at governmental level, albeit with some contact with the Assemblé[13].

At the meeting of the six ministers of foreign affairs in Paris on 12/13 May, Adenauer did not take up the position the Dutch had anticipated, but one which for them was much worse. What the ministers had before them, he argued, was a draft treaty which would enable the ECSC and the EDC to be married in a system of democratic control. And that was what the ministers should strive for without trying to give the EPC any other attributes, which would only delay or might even endanger its implementation. There were a few institutional questions to be sorted out and then it could be left to government experts to draft a final text. This extreme 'minimalist' position was not shared by Hallstein who, in Adenauer's absence, suggested that the German government could largely agree with the draft treaty and was not averse to giving the Political Community functions outside that of controlling the existing institutions as long as these were firmly agreed beforehand. However, he did share the Chancellor's concern that the treaty be agreed as soon (and therefore with as few amendments) as possible. The German position, as stated by Hallstein, was most closely approached by de Gasperi who wanted the inter-governmental conference envisaged by the EDC treaty to be called as quickly as possible so that a final treaty could be agreed. At the other extreme, however, van Zeeland pointed out that far from dropping the economic clauses in the draft treaty, as Adenauer appeared to be suggesting, it was in these very areas that the treaty should be expanded. This position was supported, naturally, by Beyen and also by Bidault who, waving the text of the draft treaty in his hand pointed out that half the articles had nothing to do with the ECSC or the EDC at all. What was involved was 'une Europe presque totale'. In the end a hopelessly optimistic timetable was agreed. There would be a further ministers' meeting in a month's time which would be immediately followed by an inter-governmental experts' conference whose work would be finished a fortnight later and ten days after that, the Ministers would meet again to review the results[14].

For France it was a useful delaying manoeuvre to set about studying all these other aspects of the problem and Bidault now even requested that these should be agreed first before the purely military and political aspects of the treaty were ratified. In one sense this made the Dutch proposals merely abstract because the strong probability now was that the treaty would not be ratified by the French national assembly and any delay in presenting it only increased this probability. But in another sense this meant that in the world after the rejection of the treaty the Dutch proposals would be on the agenda for action as the least objectionable

aspect of the treaty. Beyen obtained from the ministers agreement that a committee be set up to study the common market proposals at once and report on them to the next ministers' conference. Again this received van Zeeland's support.

The difficulties of the French government, the general elections there, the voluble opposition to the treaty, the German determination not to see a policy on which they had staked so much collapse, led to successive postponements of the alleged moment of decision until the end of September, when the full meeting of the foreign ministers and their deputies in Rome, originally set for July, eventually took place. There were less formal meetings in Paris and Baden-Baden in the interval concerned with the political and military aspects only. This long interval gave the Dutch government time to take stock. So long as Beyen's proposals had involved only statements of principle they were useful and harmless. Setting up a study committee to report to the foreign ministers, however, meant that after the débacle it was their report which might be on the agenda. What it was the Netherlands would then really want out of the situation required further definition.

Even before the Rome meeting in May the prime minister had objected that Beyen's proposals, even if written into the draft treaty, would not go far enough to meet Dutch needs. The proposals foresaw that the executive body of the EPC should fix a programme for abolishing tariffs and quotas and that this programme should be subsequently approved by the parliamentary organs. As far as Drees was concerned there must be a guarantee that something would actually happen, that there should be an automatic process of tariff and quota reduction and not one dependent on the voting in the Executive and the parliamentary assembly[15].

The difficulty here was to get agreement at international level on an automatic process. The opinions elsewhere as tested out by van der Beugel and Linthorst Homan on a tour of the capitals in May did not suggest this would be easily reached. Belgium was the only other strong supporter of the customs union concept but the standpoint in Brussels was not the same as in the Hague. A customs union would, the Belgians argued, have also to provide for freedom of capital and labour movements. It could not, however, cope with changing the monetary systems and policies because currency convertibility, which was essential to the operation, would have to be reached and managed in a wider framework than that of the six. Elsewhere, Italy and Luxembourg while not openly opposing the Dutch proposals showed little enthusiasm for them, except in the sense that if 'Little Europe' were to be created some form of economic integration would have to be part of it. In the Federal Republic opinion could hardly be weighed up, it depended on which ministry was assumed to be making policy and on guesses about the power of Adenauer to make recalcitrant ministers accept his view of the prime political importance of the treaty.

As for Monnet himself, although of course he would have accepted the widening of the ECSC into the EPC, he was strongly against the Beyen plan as a definition of the purposes of a future EPC[16]. The Ministers' meeting in Baden-Baden in August provided a further opportunity for testing the waters for the reaction of the others but it provided little by way of new insights. Taviani was slightly more positive in expressing the Italians' support than might have been expected from van der Beugel's and Linthorst Homan's report, but he warned that the EPC was a political statute and not an economic one. Bech expressed Luxembourg's reservations on the free movement of labour and the removal of agrarian protectionism, at least as far as Luxembourg was concerned. But since neither Italy nor France nor Germany had come prepared for involved economic discussions, that was as far as it got[17].

As early as the beginning of June the Dutch cabinet set about fixing the instructions of the delegates to the Rome conference. Beyen and Luns suggested that the delegates be instructed to push for a term of 10 years to be written into the treaty for the realisation of a customs union, a point which was eventually accepted despite the fact that both Drees and Zijlstra considered it too long. Beyen pointed out that if the time-scale was shortened it would lead to an increase in the number and complexity of the escape clauses. Aside from that, the opinion in cabinet was unanimous that the delegates should work from the basis of the Beyen Plan, that they should refer back to cabinet any possible compromise suggestions, and that the Netherlands should not join the EPC unless it had a real economic content[18].

The experts met together in plenary session in Rome for the first time on 22 September and confined themselves to a general exposition of national positions. The Belgians were firm in their standpoint that the Community should be more than an organ linking the ECSC and the EDC. On the institutional side, they wanted collective responsibility in the Executive (i.e. that the parliament could not insist on the dismissal of one member) and 'paritair' representation in the senate. Moreover they wanted a secession clause introduced into the Treaty. Hallstein, representing the German delegation, wanted far more clarity on the relationship between the various proposed organs of the Community. He was not opposed to the incorporation of the EDC and ECSC as envisaged in the treaty, seeing any difficulties as technical rather than political. Finally he supported the Benelux position of wanting more economic tasks entrusted to the Community, emphasising in particular the monetary question. The Italians, too, supported the Benelux position on economic questions and the German position on the question of incorporating the two communities. They supported a directly elected parliament, but wanted the membership to reflect more closely national population distribution, though they were willing to consider a 'paritair' senate. Tjara van Starkenborgh

Stachouwer, leading the Dutch delegation, stressed the Netherlands' demand for automatism in tariff integration and reduction, the reserve on incorporating the other two communities and the postponement of direct elections. The Luxembourg delegate confined his opening comments to a demand for a 'paritair' senate, whilst the French declared themselves unable to make a declaration at all at that moment[19]. The reason for the French position was broadcast on the French radio the following morning – namely a split within the government between the MRP, who wanted a construction for the Executive similar to that argued in the draft treaty, and the Gaullists[20]. At the next session, the French prime minister asked for understanding for the difficult position in which he found himself, but he was at least in a position to reject as unacceptable the Dutch demands for automatism in the procedure towards a customs union. Moreover it became clear that he wanted the economic clauses to be framed very generally so as to be subject to later treaties. On institutional questions, he considered a directly elected parliament acceptable and a 'paritair' Senate possible, but he did want a shift in the balance between the Council of Ministers and the Executive in favour of the former. To this Hallstein had replied that Germany, too, considered a political community shorn of economic clauses useful in its own right as promoting 'sachliche und politische Ziele'. With the battle lines so drawn, the experts decided to entrust their further deliberations to two working groups, one devoted to economic issues and the other to institutional questions[21].

The differences apparent in the opening sessions manifested themselves immediately in the first session of the economic working-group. At the one extreme were the French who argued that the incorporation of the EDC and ECSC into a European Community was an important step in itself and warned not to move too far ahead of public opinion in trying to do more. Closest to them, though at some distance, were the Italians who argued that the new Community should have some economic tasks but wondered if everything had to be regulated by treaty at this stage. The Dutch, hammering on the need for a timetabled customs union, found themselves in the middle-ground. Next came the Belgians, who, whilst supporting them on this question, asked if a customs union, without measures to coordinate national economic policy, was not an 'adventure'. The Germans argued that monetary and financial questions were of primary importance, but made no statement on the desirability of a customs union itself. Luxembourg, according to the Dutch report, 'made some unimportant comments'[22]. The following day was devoted to a largely inconclusive discussion on how further to proceed but it did provide three additional standpoints: that the Germans were prepared to consider a certain degree of automatism in the achievement of tariff reduction; that the French were willing to cooperate in any common investment policy to remove obstacles in the way of a common market and

to give the Community the right of initiative to make 'propositions and recommendations' for the achievement of that market, and that Luxembourg could agree with anything the rest decided as long as it did not have to surrender the protectionist privileges it enjoyed in Benelux![23].

The next day the various standpoints began to clash for the first time when the delegates settled down to discuss the issue, 'Definition of the common market'. No-one was particularly happy with the definition tabled, which was limited to goods only, and eventually 'services, capital and labour' were added. It was at this point that the Germans wanted added to the definition the statement that, 'the achievement of this final goal presupposes a financial and economic policy based on analogous principles, which would in particular have the effect of guaranteeing the financial stability of the member states'[24]. This was broadly supported by the Belgians but not by the Dutch, who argued that a common market referred solely to the removal of economic frontiers. If, as a result of an inflationary policy in one country, it ran a balance of payments deficit with the rest, the country could either draw on its reserves of international credit or it could allow its exchange rate to fall. In either case the common market would have been left intact. Thus one could not argue that political coordination of financial and economic policy was a precondition for a common market, let alone part of the definition. The French, whilst not accepting the full rigour of the Belgian/German position, argued that it should be seen as a precondition. If a country, for historical reasons, had a high price level, it would import more or export less. Unless such a country were able to limit its imports it would soon encounter great financial and monetary difficulties. The problem, explained the French, was that the ministers at Baden-Baden had said that the states must retain sovereignty, so that it was impossible to agree to such a measure of policy coordination as the Germans proposed, since that would mean, for example, that France would have to discuss its Indo-China difficulties with the Community, which was unthinkable. The whole matter, they argued, required a new *political* decision by governments after Rome[25].

The deliberations then shifted to the question of 'measures to be taken in order to achieve the common market'. The Germans had submitted an aide mémoire which placed monetary and financial harmonisation on an equal footing with the elimination of trade barriers. On the first point it stressed the need among member states for budgetary equilibrium and a monetary policy sufficiently constrained to ensure currency convertibility on current account at least. On the latter it stressed the gradual abolition of tariffs and quotas on intra-group trade and the erection of a common external trade regime 'appropriate for the encouragement of international exchanges'[26]. The discussion first centred on the point of the external tariff and surprisingly there was general agreement on the fact that it should not be autarchic, but that still left a wide range of alternatives

ranging from the lowest to the highest or a trade-weighted average per product. Presumably thankful that there was some agreement, that question was pushed aside. After the Germans assured the Dutch that it had not been their intention in stressing budgetary equilibrium to rule out the possibility of countries following a counter-cyclical policy, the session was adjourned[27]. Two major questions had been left hanging in the air, how were financial and economic policies to be coordinated and exactly what was to go into the treaty concerning the creation of the customs union? That discussion was bound to come but, in the event, it was triggered by the Italians who argued that it was inappropriate to fix executive steps and dates for a process of integration which would demand a long period of preparation and evolution[28].

The German position was defined before the meeting by von Maltzan in a private conversation with Linthorst Homan. 'The only thing we want', he said, 'is that we six should not retreat on the monetary question'. Referring, presumably, to the British proposals for a more general, coordinated approach to currency convertibility in the western world he described these proposals as 'immensely dangerous, because they could break apart our group'. In the meeting itself he proposed that the six form a 'convertibility club' which could take initiatives towards real convertibility in a group larger than the six[29].

The German solution both to the Dutch demand for 'automatism' and for the question of policy harmonisation was to give the Community rights of decision and the power to make its own legislation. This was supported by the Belgians since it would ensure that countries could not sabotage progress towards the common market. For the French, it was unacceptable to move rapidly to a Community; given over to the old liberal ideas of 'the game of competition' which would lead to the enrichment of the rich and the impoverishment of the poor. Would it not be better 'to integrate in order to liberate rather than to liberate in order to integrate'? However, the German idea of giving the Community its own legislative powers would erode national sovereignty. 'Recommendations', rather than 'decisions', were the most they were prepared to concede to the Community's Executive[30]. It was at this point that the discussions were suspended since the experts had to knock together some kind of report and there was only a week to go. Because of the lack of time, it was agreed that the report would list a number of general principles and that the position of each delegation would be recorded on each point[31].

Looking at the conference as a whole one thing was clear from the start, the French reaction to the inclusion of any commitments to 'economic integration' was negative in the extreme! This in itself handicapped the attainment of much consensus over the various elements of the Beyen Plan. A further problem was that the Beyen Plan (as it existed in its three separate memoranda) was not accepted as a conference document. More-

over, the Dutch delegation felt that they were handicapped by the very detail in which the Beyen Plan was worked out. Almost every detail called forth a psychological or political reaction and the overall direction of the proposals tended to get lost. Another factor was that there was very little negotiation in the strict sense of the word. Finally economic integration, to be successful, required both liberalisation and harmonisation but the Dutch standpoint seemed to require automatism in the first case but not in the second. Because what the Dutch wanted was a guarantee that something would indeed happen, their delegation pressed hard for room at the next foreign ministers' meeting to make concessions which might produce a better front against the French[32]. Even while the Conference of experts was going on, there were clear signs of a split beginning to emerge within the Dutch cabinet on this question. On 28 September Mansholt wrote to Beyen urging him to modify the delegates' instructions. He pointed out that the German sympathy for Dutch aims for economic integration was beginning to evaporate in favour of an acceptance of a political treaty which, in line with French thinking, would do no more than link the ECSC and the EDC within a common institutional framework. Although he himself was not against such a treaty, he felt it would hamper further progress towards economic integration. Yet this development was difficult to deflect given the instructions binding the Dutch delegation which did not allow it to make any concessions to the Germans on monetary questions or to the Italians on social issues. Such concessions would make it clear that the Dutch viewed economic integration in terms larger than a tariff union. As long as this was the case, the French had a free hand and the danger increased that the economic paragraphs would disappear from the Treaty altogether. He felt that the Netherlands should accept a situation in which a common market would be regulated by a separate agreement or else that the Community itself be instructed to prepare such an agreement. 'This limited *auto-extension* has so far been rejected by the Dutch government, but if our concrete plans should strand on too much resistance, I would prefer it above the omission of every sentence on economic integration in the draft treaty and also above a purely platonic mention in the goals'[33]. It shifted Beyen not one inch. In a somewhat high-handed reply he argued that the reaction to the Dutch proposals in Rome surprised him not in the slightest but 'it has been our strength that amid all the wavering, changes in position and opportunism we have met from our friends in this company we have stayed firm by a simple and clear proposition'. Political integration without a task in the economic field would be an 'empty husk' which would do Europe more harm than good. The purpose of the Rome Conference was to inventory government reactions and it would be premature to alter the Dutch position since it would only create uncertainty and confusion over Dutch intentions. After Rome would be the appropriate time for the cabinet to take new deci-

sions[34]. When Mansholt raised the question of introducing some suppleness into the Dutch negotiating position in cabinet, he received no support whatever. Beyen did not see the risk that the French and Germans would push for a limited treaty as very great. Moreover he argued that the minimalist French stance was more a reflection of the fact that it had *no* position (*vide* the disagreement in cabinet) than an indication of what French policy would be when it eventually emerged[35].

At the end of October Beyen produced a memorandum outlining the range of possible Dutch concessions. The target date for achieving the customs union could be set back to fifteen years. Because no one else would accept the automatism of tariff reductions it might be possible to stipulate that half the reductions be automatic and completed halfway through the transition period while the executive body could decide on the rest. It might be possible to accept Belgian and German wishes to give the Executive powers over policy harmonisation, but only if the removal of trade barriers proceeded at the same pace. This was not much. And it must not prejudice the Dutch position that steps to the customs union must not be dependent on a unanimous voting procedure in the Council of Ministers of the EPC nor upon repeated parliamentary ratifications[36].

In cabinet Beyen acknowledged the danger that the forthcoming foreign ministers' meeting in The Hague might reach agreement on the constitutional questions whilst the economic issues could be referred for further study. This would be completely unacceptable to Drees who felt that you could only establish the form of the EPC once it was clear which tasks it should get. If agreement were reached on the political points the Dutch would be under pressure from the others to let it operate immediately and leave the solution of the economic difficulties till later. He felt that there should be no compromise on the political front until agreement on economic questions was reached, a point which was supported by other ministers. Moreover there was a majority in cabinet against making the time-span for the implementation negotiable[37]. Such, then, was the position of the Dutch government on entering the The Hague meeting - a position which, in practical terms, boiled down to pushing for further study of the EPC in its entirety by groups of experts[38].

At the last moment Mansholt tried again to convince cabinet that it was essential to bring the EPC into operation quickly. This meant abandoning its standpoint on a customs union, giving the task of the realisation of economic integration to the proposed executive body and recognising that nations had a right of veto. This found virtually no support. The views of Drees were representative, 'Even if the French only want a head-dress for the ECSC and the EDC, the Netherlands should still say no, in view of the fact that we should not be prepared to surrender part of our sovereignty for the creation of an empty husk'[39].

Even this was probably overstating French ambitions by 26 November

when the Hague meeting of foreign ministers took place. Bidault himself could not attend until the last day and that, together with the fact that the government had not been able to discuss its position with the French parliament meant that the meeting was unlikely to get very far in making decisions (which for the Dutch/Benelux strategy was possibly an advantage since it made it easier to push the case for the installation of a study commission). In fact the proposal to set up a study commission was accepted relatively painlessly (though one could wonder why the French should bother since Parodi explained that as far as the French were concerned the fundamental aim of the EPC was to link the ECSC and the EDC under a single form of democratic control). The basis for the group's work was to be the report of the Rome conference and it should have its report completed by 15 March 1954[40]. The cabinet decided that no new instructions for the delegates to the Paris study group were necessary[41].

It hardly seemed any longer a matter of real concern. The Study conference did not meet until 8 December and the negotiations proper did not start until 7 January. It then divided into three commissions – an economic commission, an institutional commission and a commission to study the question of European elections. Of these, the economic committee was the most important for the realisation of Dutch goals but it took almost a week to get going at all largely because the French spokesman, Wormser, kept insisting that the discussions should not go beyond the terms of article 38 of the proposed treaty and that if the Community were to assume new economic tasks, that was something over which it could deliberate and upon which governments could decide at a later date. Eventually, to stop the talks stalling altogether, it was agreed that French reserves could be stipulated and recorded on every single sentence if necessary[42]. However, although the schism separating the French from the rest was the most evident, and from the Dutch standpoint the most welcome, there was a further split which was much less promising: namely that Italy and Germany seemed to be thinking more in terms of a *traité normatif* rather than the *traité executif* which the Benelux countries considered necessary to ensure some real economic progress[43]. The Economic Committee, however, did not so much function as a 'study-group' or a 'negotiating platform' at all in the early weeks but attempted instead to unravel the text of its report at the end of the Rome conference and to stick it back together again in a somewhat more logical fashion. The one positive point was that the Germans pulled back from their flirtation with a *traité normatif*. On the other hand, the Benelux cooperation which had been set up before the Paris talks, appeared to be paper thin. In the first place, in the first days, the Belgians had presented a list of questions to form the Committee's approach, without prior consultation with the Dutch, which did not appear to give primacy to a free exchange of goods as a principle goal (in the event, nobody paid any intention to the list

anyway). More damaging to mutual cooperation was the Belgian attack on the Dutch system of *sauvegarde* suggesting that it would be rendered totally unnecessary if liberalization were preceded by a system of strong policy coordination. Not only did this go beyond earlier Belgian sentiments on giving the executive powers to impel monetary coordination, but it propelled the Belgians almost directly into the arms of the French who wanted nothing better than *intégrer pour libérer*[44].

The French were distantly tolerant of the rest wanting to concretise points which they considered irrelevant whilst the Italian delegation still had received no instructions owing to a simmering cabinet crisis[45]. One of the first points was the realisation of the customs union, or was it an economic union or was it something in between? By the time they had sorted that out, what was to become the new French strategy made its first appearance. Should not the appeal to GATT be made on the basis of article 25 instead of article 24 since certainly clauses 5-c and 7-c of article 24 were less attractive than a possibly revised article 25. Of course France's reserve remained intact that, as far as they were concerned, none of this belonged in the treaty. After the delegates had followed that red herring to the bitter end, the Germans said that they were not sure it was relevant anyway since they at least had not agreed to the automatism in the Dutch proposals and that they could not really talk about GATT until the question of harmonisation had been discussed. At this point Linthorst Homan seems to have lost his temper and suggested that if the others did not like the Dutch proposals, why did they not come up with some concrete alternatives themselves so that there was at least something to discuss. When the discussion resumed only Belgium supported the Dutch position that an end-term for a customs union should be included in the treaty. Germany suggested a mixture of automatism and supranationality whilst the Italians felt that the whole question depended upon the degree of progress towards policy harmonisation and the removal of 'structural barriers'. The new German suggestion lay in the direction of the compromise Beyen had already considered making and was, moreover, acceptable to the Belgians[46]. The Dutch, however, did not make that concession and so, when it came to discussing how tariffs could be lowered, the gulf opened up again[47]. It was at this point that the French made a proposal for a preference agreement – mentioning specifically, as an example, French grain exports. The Dutch chairman referred to this as a 'bomb in the conference hall' to which the French replied that it was a normal extension of European solidarity which could be decisive for French public opinion. The question was shelved[48].

On the question of free movement of labour, the Germans and Belgians expressed the worry that the ethnic composition of certain regions could be affected; a point conceded by the Italians who felt that safeguarding clauses could play a useful role. Germany, moreover, ex-

pressed reserves about whether the principle should be extended to cover all professions (doctors, lawyers etc.) but agreed to look into the matter again. Luxembourg wanted itself exempted from whatever the rest agreed.[49]

Policy coordination was also discussed by the experts for the first time in any detail. Germany emphasised the distinction between positive coordination which would help promote changes in economic structure and thus facilitate the realisation of a common market, and negative coordination in the sense of removing differences in national policies or regimes impeding its realisation. Belgium also enumerated the circumstances which might warrant community intervention – monetary and financial instability, unfair competition and balance-of-payments disequilibrium within the six. When the Dutch suggested that an opening be left for the Community to intervene in areas not foreseen, the Belgians rejected it on the grounds that this would imply auto-extension. The Community, according to the Belgian suggestions, should be able to give directives on questions of discount rates, bank cash and liquidity ratios, the level of central bank loans to governments, the development of the national debt and the size of budget deficits[50]. There followed a discussion on the issue of safeguarding clauses which was now more or less the mirror image of the co-ordination discussion – those who wanted the first saw little need in the second and vice-versa[51]. And then everyone settled down again to prepare an agreed text which properly reflected the full extent of disagreement.

At this point the French reverted to a minimalist/obstructionist position which, the Dutch observed, only served to drive the rest closer together. Moreover, the German tendency not to push too hard for their own position and to try to omit areas from discussion which might prove unacceptable to the French was now abandoned. Even the Italians who, in their desire for a *traité normatif* had been closest to the French position were now completely 'uncoupled'. The report itself had resolved very little. The time had come for *political* decisions[52].

Two months of work in Paris had altered very little when compared with the situation arrived at in Rome. There were a number of reasons for this. The first was that the Rome report was declared by the Hague Meeting an 'official' document, even though it reflected little more than the listing of six sets of government instructions to the respective delegates. Although the Paris conference was supposed to be an exchange of views by unprejudiced experts, it appeared that in Paris, too, they were tied by government instructions and that any divergence from the Rome Report was seen as a sign of weakness. A second factor was the approaches of the delegates themselves. The French delegation, the Dutch reported back, 'continually attempted to throw sand into the works by continually trying to initiate discussions on problems of detail; the Germans are extremely cautious and hardly ever speak out clearly or else they try to

push problems aside as a result of which they often continue to travel round in a vicious circle of vague slogans; the Italians seem to be working on direct instruction from Rome and refer back every important point and the Luxembourgers, true to tradition, remain silent'. But the Dutch themselves were equally to blame as Linthorst Homan (who was heading the delegation in the economic commission) pointed out caustically. Their freedom for manoeuvre,outside the minor concessions authorised, was minimal. They had detailed proposals in one area (which was seen by all the other delegates as insufficient) but no room in their instructions to be able to consider seriously the proposals of the others on other areas. This in fact could have wasted the chance of forming a closer position with Belgium, Luxembourg, Germany and Italy none of which was against trade integration but saw the problem of a 'common market' in a wider perspective than the Dutch but the Dutch could make no concessions. In the end Van Starkenbourg Stachouwer (head of the overall Dutch delegation) had brought this to Beyen's attention and in drafting the final report, the Dutch had dropped their reserves on a number of points concerning policy coordination. However, not only was this too late in the day to make any difference, but, as Beyen cynically explained afterwards, it would demonstrate that the Dutch at least had seen the exercise as a study which did not bind the governments; he could always reintroduce Dutch reserves at a later date. That was about the only movement achieved in the two months in Paris[53].

When the Dutch cabinet met to review the results of the Paris study conference in April 1954 the clearest point visible to all was the gulf separating France from the rest. Louis Beel, the Vice Prime Minister, commented on the irony that the Dutch had gone along with the EPC because France apparently wanted it as a condition for ratifying the EDC. The situation now was that France seemed to be exceptionally negative on the question of both the EPC institutions and the powers it was willing to give them and that the ratification of the EDC seemed to have split the country down the middle. The general consensus in cabinet was that there was little point in formulating the Dutch position until it was clear what the French were going to do. Mansholt, who thought that the Dutch should push ahead allowing the new EPC organs to implement the proposed 'economic integration', found himself in a clear minority[54].

The next Council of Ministers meeting asked the study conference to resume its work. So in May the experts trudged back again to Paris to resume their work and it was decided in both committees that the aspects they would study should be as 'neutral' as possible. In the economic committee the French delegate, Soutou, had explained that his country's position had been so vague in the earlier talks because of French difficulties in the OEEC. Since these had been resolved, a new dynamism had been imparted to French economic policy and they could now talk more

openly. 'We are emerging from the circle of protectionism', he proclaimed[55]. Whether any of the others believed him is not clear. When the talks began in earnest, however, it became apparent that little had changed. The French position remained that everything would be settled in a later treaty whilst the Germans had reverted to a conciliatory stance towards France though, rumours had it, that this was partly the result of a struggle between the economic and agricultural ministries over the economic implications of freeing trade[56].

Nonetheless a number of new ideas did emerge. On the question of how to implement a customs union the Germans revived the idea that a starting period of two or three years should follow the working of the treaty, after which the Community would proceed to the implementation of a customs union. This band-wagon was immediately jumped on by the French who had always wanted a starting period, although a much longer one, before moving towards integration. Since, by this stage, it was unclear who was talking about what, the others stuck to the position that the date for the implementation of the customs union had to be fixed in the treaty[57]. As for the question of how to arrive at that customs union, the Germans outlined a number of routes other than the strictly arithmetical approach of the Beyen-Plan: differential tempos for different sectors, reductions by weighted averages, beginning by 'capping' highest tariffs and exempting the lowest and then proceeding to a common scheme. The Dutch now moved away from the Beyen scheme and declared a willingness to consider any scheme as long as it included a measure of automatism and ended, at a specified time, at a nil-tariff on intra-group trade, but that was still insufficient to bring round the French[58]. As far as the external tariff was concerned the range of options lay between the French insistence that it should be high, so as to improve the chances of GATT concessions, and the Benelux position that it should be as low as possible. This succeeded in resurrecting the French idea of preferential purchasing – if, for example, the duty on grain were low, there would have to be an agreed grain-programme. The position of the Germans and Italians did not emerge[59]. At the last meeting, in July, the discussion focused primarily on a Belgian document on Benelux. It was agreed to put together a report and to meet again 'in the autumn'.[60]

In August Mendès-France called together a meeting of the Six governments which met in Brussels, at which France demanded a number of further concessions on the EDC Treaty which, he argued, were necessary if it were to have a chance of success in the French parliament. The other five were equally resolute in their refusal to amend a treaty they had already ratified, especially since the French were not willing to put anything in its place[61]. On 31 August the French national assembly rejected the EDC. Since the EPC existed solely by virtue of the EDC, the whole edifice came tumbling down. The whole question of German rear-

mament was eventually resolved by the creation of the Western European Union, but the push towards 'economic integration' seemed to have come to a grinding halt.

Looking back over the economic negotiations within the context of the European Political Community it is important to bear in mind that they all took place at a time when few political commentators would have given the European Defence Community, which underpinned the whole structure, even the slightest chance of ever being ratified by the French Parliament. Yet armies of diplomats, civil servants and so-called experts from six countries were locked together for months in Rome and Paris discussing the economic and political structure of a new community. The archive situation allows us to do no more than speculate over the answer to this apparent paradox. The first possibility is that the leaders of the six were trapped by the very expectations which they had created and that none of them was willing to attract the odium for allowing the initiative to fail. This seems highly improbable, certainly as a consensus, partly because it seems likely that at least some of the six had other motives for keeping the initiative going and partly because, if such had been the consensus, the best tactic would have been to allow matters to drift and not to begin by setting up layer upon layer of study groups. The French position of frank hostility towards the study groups was certainly consistent with this second attitude. But if it was not a dance of the dead, what was it? The second possibility lies at the other extreme of the spectrum namely that leaders such as Adenauer believed to the bitter end that the EDC would be created and that he, for example, was supported in that by the American Secretary of State, Dulles. Along this line of thinking, once the log-jam had been broken, the EDC, the EPC and a limited economic community would all, in turn, come into existence. However, this alone would appear to be an insufficient explanation. The explanation becomes more plausible when a third possibility is admitted namely that, even if the EDC were to be written off as dead, there were those European federalists who considered that the EPC itself was worth salvaging. Certainly the Italian position, which throughout the whole episode remained closest to the original conception of the draft treaty of the *Assemblé ad hoc*, would suggest that de Gasperi considered this a real possibility. Add to this a fourth option, that the Dutch and the Belgians, whilst disagreeing over the best approach, considered it important to keep the issue of European protectionism on the political agenda even if progress were minimal, and that there was a majority which for diverse reasons had an interest in keeping the political process alive.

For the Dutch at least the failure of the EPC had temporarily cut off one route but the tariff question remained unresolved and its resolution would ultimately have to come from an international solution. Eventually, it could be argued, a situation would arise when other countries wanted

something else so much that they would be willing to pay the Dutch price for its participation, in the form of an agreement to lower tariffs. After all, the Dutch institutional reservations had all been accepted early on in the Schuman plan negotiations because France and Germany wanted a coal and steel pool. In the same way, the Beyen Plan had only got onto the agenda in the first place because France, Germany and Italy had at the time wanted a form of democratic control over the EDC. Thus, when a year later the Beyen/Spaak initiative linked the call for the creation of a customs union with a plan for the joint regulation of atomic energy, it could be interpreted as an astute move to create a situation in which the Dutch bargaining position could be maximalised.[62]

## Notes

[1] 1. Algemeen Rijksarchief, The Hague, Ministerraad Archive. ARA, MR (482) *Nota aan de Ministerraad inzake Ministers conferentie 9 en 10 September te Luxembourg* (MR 5.9.1952).

[2] ARA, MR (482) *Kort verslag van de Eerste Zitting van de Bijzondere Raad van Ministers van de Europese Gemeenschap voor Kolen en Staal en van de daarop aansluitende Conferentie der Ministers van Buitenlandse Zaken van de staten-leden dezer Gemeenschap, gehouden op 8, 9 en 10 September 1952* (MR 15.9.1952). See also J.W. Beyen, *Het Spel en de Knikkers. Een kroniek van vijftig jaren*, Rotterdam 1968, 205-209.

[3] ARA, MR (483) *Nota der Advies Commissie inz. de instelling ener Europese Politieke Gemeenschap, betreffende het van de zijde der Nederlandse Regeering in te nemen standpunt dienaangaande, en in het bijzonder omtrent de aan de Assemblé ad hoc voor te leggen vragen* (MR 16.10.1952).

[4] ARA, MR (484) *Grondslagen voor het Nederlandse standpunt met betrekking tot het vraagstuk der Europese integratie* (MR 17.11.1952).

[5] Beyen, *Het Spel en de Knikkers* 227-229. ARA, MR (486) *Ontwerp van een Memorandum bestemd voor de regeringen* dd. 28.1.1953, *Europese Politieke Gemeenschap* dd. (MR 9.2.1953).

[6] ARA, MR (487) *Verslag van de Ministers Conferentie van de zes Schuman landen te Rome op 24 en 25 Februari 1953* (MR 2.3.1953). Min. BZ, 913.100/23 *Reacties van de verschillende Ministers op de Nederlandse voorstellen* dd. 7.3.1953.

[7] ARA, MR (488) *Ontwerp Verdrag behelzende het Statuut van de Europese Gemeenschap* (MR 29.4.1953).

[8] Archive of the Ministerie van Buitenlandse Zaken, The Hague. Min. BZ, 913.100/26 Council of Europe. Consultative Assembly. 5th Ordinary Session, May 1953. Recommendation 45 Appendix II.

[9] Ibid., *Rapport van der Economische Sub-Commissie* dd. 22.4.1953.

[10] Ibid., *Staatsrechtelijk-Institutionele Beschouwingen omtrent het ontwerp-verdrag inzake de Europese Gemeenschap* dd. 23.4.1953.

[11] ARA, MR (398) Minutes of Cabinet 29.4.1953.

[12] Min. BZ, 913.100/25 *Concept voor Nederlands Memorandum inzake de Europese Gemeenschap*, n.d. Accompanying letter dated 4.5.1953. ARA, MR (489) *Memorandum du Gouvernement des Pays Bas concernant la Communauté Européenne* (MR 11.5.1953).

[13] ARA, MR (398) Minutes of Cabinet 11.5.1953. See also ARA, MR (489) *Projet de Dispositions économiques du Traité portant Statut de la Communauté européenne* n.d. (MR 11.5.1953).

[14] ARA, MR (489) *Verslag van de op 12 en 13 Mei gehouden Ministersconferentie betreffende de Europese Gemeenschap* (MR 26.5.1953).
[15] ARA, MR (398) Minutes of Cabinet 11.5.1953.
[16] Min. BZ, 913.100/27 *De eerste reacties op de Nederlandse voorstellen* dd. 30.5.1953.
[17] ARA, MR (492) *Verslag van de vergadering van de zes Ministers van de KSG-landen, gehouden te Baden-Baden op 7 en 8 Augustus 1953* (MR 24.8.1953).
[18] ARA, MR (398) Minutes of Cabinet 20.7.1953. ARA, MR (491) *Instructie voor de Nederlandse Delegatie, welke zal deelnemen aan de Intergouvernmentele Conferentie voor de Europese Gemeenschap* n.d. Covering letter dated 12.6.1953 (MR 20.7.1953).
[19] Min. BZ, 913.100/13 Code telegram van Starkenborgh to BZ, 28.9.1953.
[20] Ibid., *Meningsverschillen in Franse Regering over instructie aan de Franse delegatie bij besprekingen te Rome over de Europese Politieke Gemeenschap*, dd. 23.9.1953.
[21] Ibid., Code telegram van Starkenbourg to BZ, 24.9.1953.
[22] Ibid., *Kort Verslag Economische Werkgroep 23 September* n.d.
[23] Ibid., *Kort verslag 24 September 1953.*
[24] Ibid., *Kort verslag Economische werkgroep 25 September 1953, Commission Economique. Procès-Verbal de la troisième séance* CIR/CE/PV3. This was the first located set of semi-official minutes.
[25] Min. BZ, 913.100/31 *Kort verslag Economische Werkgroep 26 September, Commission Economique. Déclaration de la délégation néerlandaise sur la relation entre le marché commun et la coordination de la politique monétaire* CIR/CE/Doc 4. The semi-official minutes *Commission Economique. Procès-Verbal de la quatrième séance* CIR/CE/PV4 did not record the exchange.
[26] Min. BZ, 913.100/30 *Commission Economique. Aide Mémoire de la Délégation Allemande sur les mesures à prendre pour réaliser le marché commun* CIR/CE/Doc 5.
[27] Min. BZ, 913.100/31 *Kort verslag van de Economische Werkgroep 28 September, Commission Economique. Procès-Verbal de la cinquième session* CIR/CE/PV5. *Déclaration de la délégation néerlandaise sur les questions formulées par la délégation allemande concernant les mesures à prendre pour réaliser le marché commun* dd 28.9.1953.
[28] Ibid., *Commission Economique. Déclaration du Délégué italien concernant les mesures à prendre pour réaliser le marché commun* CIR/CE/Doc 7.
[29] Ibid., *Kort verslag economische werkgroep 29 September 1953.*
[30] Ibid., *Kort Verslag Economische Werkgroep 30 September 1953.*
[31] Ibid., *Economische Werkgroep 2-10-'53 Vergadering Redactie-Commissie II.* For the results see Commission Economique. *Rapport du Comité de Direction* CIR/CE/Doc 10 and *Paragraph VII revisé du Rapport au Comité de Direction* CIT/CE/Doc 18.
[32] ARA, MR (494) *Rapport aux Ministres des Affaires Etrangères* CIR/15, *Delegatie-Verslag van Conferentie Europese Gemeenschap te Rome* dd. 10.10.1953 (MR 12.10.1953).
[33] Archive of the Ministerie van Algemene Zaken, The Hague. Min. AZ, 351.88(4)075:32 Mansholt to Beyen 28.9.1953.
[34] Ibid., Beyen to Mansholt 3.10.1953.
[35] ARA, MR (398) Minutes of Cabinet 5.10.1953.
[36] ARA, MR (495) *Resultaten van de Conferentie van Rome, betreffende de Europese Gemeenschap* dd. 22.10.1953 (MR 21/23.11.1953).
[37] ARA, MR (398) Minutes of Cabinet dd. 2.11.1953.
[38] ARA, MR (495) *Door Nederland op de Conferentie te Den Haag van 26 November in te nemen standpunt* (MR 17.11.1953) *Nota betreft: Het Nederlandse standpunt op de Ministersconferentie van 26 November* (MR 17.11.1953).
[39] ARA, MR (398) Minutes of Cabinet dd. 23.11.1953.
[40] ARA, MR (495) *Conferentie van Ministers van Buitenlandse Zaken gehouden te Den Haag van 26 t/m 28 November 1953* (MR 30.11.1953).

[41] ARA, MR (399) Minutes of Cabinet dd. 30.1.1954.
[42] Min. BZ, 913.100/34 *Weekverslag van de Nederlandse Delegatie in de Commissie voor een Europese Politieke Gemeenschap (periode 7 t/m 8 Januari), Economische Commissie. Interim-commissie No. 3* dd. 13.1.1954.
[43] Min. BZ, 913.100/36 *Overzicht van de Economische Commissie over de periode 7-15 Januari.*
[44] See Min. BZ, 913.100/34 Linthorst Homan to van der Beugel, 16.1.1954, Min. BZ, DGEM 6111/1252 Kupers to van der Beugel 23.1.1954, Min. BZ, 913.100/34 *Enige hoofdpunten van de Economische Commissie te Parijs* dd. 2.2.1954.
[45] Min. BZ, 913.100/36 *Regeling der werkzaamheden, 8 Februari 1954.*
[46] Ibid., Report dd. 10.2.1954 EC 10.
[47] Ibid., Report dd. 11.2.1953 (sic) EC 11.
[48] Min. BZ, 913.100/37 *Weekverslag van de Nederlandse Delegatie in de Commissie voor een Europese Politieke Gemeenschap (periode 13 t/m 20 Februari '54)N. See also Réalisation et maintien du marché commun* dd.16.2.1954 CE/Doc.Trav. 13.
[49] Ibid., Report d. 19.2.1954 EC 15.
[50] Ibidem.
[51] Ibid., Report dd. 20.2.1954 EC 14.
[52] Min. BZ, 913.100/38 *Weekverslag van de Nederlandse Delegatie in de Commissie voor een Europese Politieke Gemeenschap (periode 25 Febr-6 Maart '54.*
[53] This summary has been put together on the basis of the following documents ARA, MR (500) *Overzicht besprekingen studiecommissie EPG, Parijs 12 December-8 Maart 1954* (MR 13.4.1954) Min. BZ, 913.100/39 *Algemene beschouwing over de economische besprekingen te Parijs studie conferentie EPG van 7 januari tot 8 Maart* (Linthorst Homan). This is much more negative about the Dutch position than the official BZ report) Min. BZ, 913.100/37 *Studieconferentie Europese Gemeenschap te Parijs* Min. BZ, 913.100/39 *Weekverslag van de Nederlandse Delegatie in de Commissie voor Een Europese Politieke Gemeenschap (periode 13 t/m 20 Februari 1954).*
[54] ARA, MR (399) Minutes of Cabinet dd. 13.4.1954.
[55] Min. BZ, 913.100/41 *Kort verslag van het besprokene in de Economische Commissie EPG te Parijs op 18 mei 1954.*
[56] Ibid., Report dd. 21.5.1954.
[57] Ibidem and *Comité Economique Compte-Rendu des travaux effectués par le Comité au cours de la période du 12 mai au 6 juillet 1954.*
[58] Ibidem and *Vergadering Economische Commissie EPG op 3 en 4 Juni 1954.*
[59] Ibidem.
[60] Ibid., Vergadering 5-6 Juli 1954.
[61] Beyen, *Het Spel en de Knikkers*, 230-233.
[62] The authors would like to extend their gratitude to the Research Council of the European University Institute for their financial support to the research on which this article was based.

# Le monde des ombres

DANIEL ROCHE
Paris/Florence

Dans la société ancienne, le théâtre et la vie n'avaient pas de frontières précises. Les balbutiements des spectacles de foire, les grandes représentations religieuses, les cortèges expiatoires et propitiatoires, les *Entrées royales*, voire les motions et les révoltes mettaient en jeu une mise en marge de la réalité quotidienne et une esthétique, une répartition symbolique des rôles sociaux et une différenciation de l'espace déjà théâtralisée. Processions, mystères, mouvements de foules ont leurs scènes. Ce que précise l'époque classique, c'est la netteté du partage dans la définition de la participation individuelle ou collective et dans l'attribution des lieux. Paysans et citadins, bien sûr, ne vivent pas cette séparation au même rythme.

## Spectacle ouvert, spectacle clos

Pour les premiers, les occasions de spectacle sont rares, toute fête religieuse ordinaire peut garder sa part de magie et devenir cet *immense conclave*, dont parle Mauss, à propos des forces collectives du sorcier. On le voit bien dans les manifestations spectaculaires des Missions; plus souvent, le paysan ne connaît du théâtre que les survivances des mystères religieux, longtemps maintenues dans les montagnes dauphinoises, les pastorales de Noël, les dramaturgies religieuses qu'interprètent des troupes ambulantes et souvent en patois. Ces comédiens des champs ne sont que de passage; des grands seigneurs, des gouverneurs de Provence – Conti en Languedoc –, des hobereaux, des nouveaux riches les attirent dans leur château où ils peuvent faire leurs affaires mieux que dans les bourgades et les hameaux perdus. Pour le XVIIe siècle, G. Mongrédien en a dénombré plus de deux cents et la liste n'est pas close: peut-être un millier de comédiens français sans compter les étrangers, italiens ou anglais, les danseurs, les bateleurs, les chanteurs. Les troupes sont constituées en société temporaire, avec un directeur (Molière n'est que le plus célèbre), regroupant une dizaine de personnes, hommes et femmes qu'unissent la

parenté et les relations amicales; le théâtre est, comme la monarchie, affaire de dynastie. Ces acteurs errants jouent où ils peuvent, salles d'auberge, jeux de paume, granges, ou en plein air sur la place du village, enfin dans le salon du seigneur. L'instabilité et la précarité sont le sort des troupes errantes sans protecteur; les autres sont pendant quelque temps mieux payées, bien vêtues, plus cohérentes car disciplinées sous le service du mécène, à son château, l'été à la campagne, à son hôtel en ville l'hiver. Pour les ruraux, le hasard fait le spectacle, ou le déplacement à la ville et à la foire, quand se rassemblent bateleurs et troupes ambulants. De leurs impressions on ne sait rien.

C'est en ville et pour les citadins que se jouent vraiment le sort du théâtre et son avenir. On voit disparaître peu à peu l'esthétique vivante qui enracinait l'expérience théâtrale du Moyen Age – principalement le Mystère de la Passion – dans des formes de participation collective et unanime, visant à proclamer l'harmonie et la solidarité de toute une société, à l'initiative des notables mais avec le concours de tous. La conjuration de l'Église et des parlements, hostiles à l'extension progressive du spectacle et à son débordement (il attire les foules venues de loin et s'accompagne de spectacles populaires que les clercs dénoncent comme des occasions de débauches, nuisibles au respect de la religion), y met fin. Dans la mesure où le profane nuit à l'édification du sacré, voire même l'éclipse, dans la mesure où l'unité sociale et culturelle qui supposait la mise en scène et la mise en rôle, puisque clercs et laïcs se partageaient les interprétations, se rompent, les vieux mystères alors quittent le devant de la scène. La moralisation des peuples et l'évolution de la culture urbaine en sont responsables. L'arrêt du parlement de Paris qui interdit *le jeu* en 1548 ne fut pas toujours respecté; en 1597, Henri IV autorise encore la représentation des mystères et la cour renouvelle son interdiction en 1598; jusqu'au début du XVII$^{e}$ siècle dans les grandes villes, et surtout dans les provinces reculées jusqu'au XVIII$^{e}$ siècle, on a joué et on a imprimé des *Mystères* épurés et surveillés qui ne favorisent plus *les rencontres et les activités scandaleuses et impudiques*. Ni le mélange des genres ni celui des publics ne conviennent à la cité classique.

La séparation qui s'instaure entre le XVI$^{e}$ et le XVII$^{e}$ siècle entre la culture des notables et celle de la *place publique* met en valeur deux espaces clos: celui des clercs et des collèges, celui du théâtre fermé où jouent les troupes fixées à Paris ou itinérantes en province. La réalité de l'exclusion du peuple s'y joue différemment. Dans la vie scolaire, on l'a déjà évoqué, le théâtre s'est imposé malgré la méfiance des jésuites: la *Ratio studiorum* essaye d'en limiter la pratique, en précisant que les représentations doivent être exceptionnelles, données en latin, et les rôles de femmes prohibés. Dans les faits, le spectacle du collège est un succès et ces injonctions prudentes ne sont pas respectées. Toutes les congrégations enseignantes introduisent le théâtre dans leur collège, on en a rappelé la vertu pédagogi-

que. C'est une des voies majeures de la pédagogie de l'émulation et pour les réformateurs scolaires, plus particulièrement pour la compagnie de Jésus, un efficace moyen de séduction de l'élite sociale et culturelle des villes. La multiplication des spectacles, le choix des dates de représentation, qui coïncident avec celles des fêtes officielles de la cité, la constitution d'un répertoire adapté et la mobilisation du public des parents d'élèves et de la bonne société, l'invitation des corps notables, municipalités, chapitres, cours, académies, en ont fait une arme remarquable de propagande et de conquête.

Dans les cours ou dans les *salles des actions* se pressent le monde des autorités, les représentants de la noblesse, les officiers et leurs épouses, les plus riches marchands, les *honnêtes gens*, les pères et mères des collégiens qui recrutent dans tous les milieux, voire même des gens du menu peuple, aux portes, et parfois entassés dans des salles contigües, tous attirés par l'éclat d'une manifestation tout à la gloire des personnages importants de la cité et à sa célébration de la hiérarchie sociale porteuse de la victoire catholique, de l'affirmation de la Contre-Réforme, de l'ordre monarchique. Sous le voile de la fable, les allusions sont transparentes; ainsi, dans les *Travaux d'Hercule* représentés à *Louis le Grand* en 1686, Hercule (Louis XIV) ruine la ville de Troie qui a violé sa parole envers les dieux (les temple abattus), aide Atlas à porter le ciel (la religion catholique soutenue), détruit l'Hydre (révocation de l'édit de Nantes). Une vision politique se joue sur scène, une dramaturgie assure en permanence, là où il n'y a pas d'autres théâtres, la diffusion de processus unificateurs. De surcroît, le théâtre scolaire a souvent donné aux auteurs profanes des sujets et un esprit; c'est à lui que le théâtre classique et ses spectateurs doivent le goût des spectacles historiques, politiques, idéologiques. C'est le répertoire collégial qui a habitué tout le monde à un cadre historique providentiel, où cependant se joue le drame du libre arbitre. C'est encore lui qui a préparé le triomphe de la tragédie classique et celui de l'affirmation des règles d'unité et de bienséance. A la fin du XVIIe et au XVIIIe siècle, c'est le théâtre de collège qui maintient le jeu des spectacles et des tragédies religieuses dans les provinces où les traditions pèsent plus que les préjugés du monde et de la ville, voire même que les objections des théologiens. La limite de ce théâtre, c'est certes qu'il est resté trop longtemps fidèle au latin, mais plus encore, c'est que, en s'identifiant à une doctrine, il se condamnait sur le plan esthétique.

Le deuxième lieu où s'est jouée la transformation du sens de la rencontre théâtrale est le *théâtre clos*: un type de scène dominé par une perspective illusionniste en profondeur, avec un public hiérarchisé de la scène à l'amphithéâtre, de l'orchestre au parterre. Cette disposition et ce dispositif scénique s'imposent à la France sur le modèle italien en même temps qu'une doctrine de la représentation dramatique; elle le fait en éliminant les autres expériences collectives ou en les déclamant. Désormais, entre les

circuits des représentations d'orientation plus élitiste et celui des dramaturgies plus populaires, l'écart va grandissant, non sans rivalité ou osmose.

## Les spectacles de Paris

C'est à Paris, capitale du théâtre, que s'est jouée la novation décisive. Dès le premier quart du XVIIe siècle, deux troupes s'implantent au temps où triomphe la *tragédie comédie*: (H. Carrigton Lancaster), celle des comédiens du Roi, à l'Hôtel de Bourgogne en 1628, celle du Marais, en 1630. Mais d'autres compagnies circulent, italiennes, anglaises, espagnoles. C'est la faveur du Roi qui fait le triomphe des *grands comédiens*, le prestige de Rotrou et la qualité de quelques acteurs; Montfleury et la Champaneslé, Beau-Château et la Du Parc. Jusqu'à la fondation de la *Comédie française*, en 1680, ils gardent la vedette sur tous les concurrents qui passent ainsi au Palais royal, Molière dès 1658, et les italiens de Scaramouche en 1660. A la mort de Molière, qui avait régné sur les fastes baroques du Versailles naissant, un regroupement s'opère par fusion de sa troupe avec le théâtre du Marais. En 1673, ainsi, est fondé le théâtre Guénégaud, qu'aménage au jeu de paume de la Bouteille le marquis de Sourdis puis, par ordre du Roi, le théâtre Guénégaud fusionne avec l'Hôtel de Bourgogne. Pour un siècle, avec l'*Opéra* qui mêle tous les spectacles, la *Comédie française* a le monopole de la scène, qu'elle étend en 1697 avec l'expulsion des *Italiens*, et qu'elle défend, à prix coûtant, contre toutes les autres initiatives, et avec succès jusqu'en 1762. Alors, les théâtres privés de la Foire et du boulevard auront enfin acquis le droit de jouer et de se développer.

Au cours du XVIIe siècle, le public s'est transformé, le changement du répertoire, le prix des places sont à la fois les raisons et les indices de cette mutation. Le *peuple* n'a pas été totalement chassé du théâtre, il compose une partie de ce public qui constitue, à la fin du XVIIe, les 100 000 ou 200 000 entrées annuelles de la *Comédie*, ce qui, compte tenu des entrées multipliées, indique une minorité recrutée à la ville et à la cour. En moyenne, le nombre de spectateurs ne dépasse pas 400 par représentation. Ils se répartissent entre les loges et les sièges sur la scène réservés à une élite du rang et de la fortune, et le parterre, où la clientèle doit rester debout. Beaucoup d'habitués ne paient pas, malgré les protestations des comédiens qui défendent la recette. Les autres acceptent de débourser vers 1715 d'une livre à une livre et seize sols selon l'emplacement, et les prix ne changent guère de 1716 à 1782. La place de parterre à un franc, c'est à peu près le revenu quotidien d'un ouvrier: le théâtre n'est plus fait pour le peuple des compagnons et des salariés, il accueille essentiellement les riches, les bourgeois aisés et cultivés, mais il s'ouvre largement, les jours chômés et le dimanche, qui voient envahir le parterre par la foule des

boutiquiers, des artisans, et la *populace*; ces jours-là, la *bonne compagnie* ne vient pas.

Les salles et leur discipline se sont adaptées à ces nouveautés. L'effervescence est progressivement contrôlée, que suscitait le mélange des classes, des oisifs et des travailleurs, des habitués qui reviennent toujours et font le succès d'une troupe et des amateurs occasionnels, des valets, des pages, des mousquetaires, chevaux légers, compagnons chômeurs, maîtres des jurandes en ballade, étudiants et collégiens, commis de boutique qui se faufilent dans l'affluence et parmi lesquels se mêlent coupe-bourses, tire-laines, filles de petite vertu. Ce public agité, plus intéressé par la farce que par la tragédie, s'interpellait, criblait les autres spectateurs de quolibets et les acteurs de trognons de choux, créant parfois de vraies bagarres auxquelles mettaient fin le *Guet* et la *Garde*: il fallait y mettre bon ordre. On riait au théâtre, et de tout, on y jouait aux cartes et aux dés, la rumeur et le bavardage étaient le fond sonore d'un spectacle qui se déroulait aussi bien dans la salle que sur la scène. Jusqu'à la fin de l'Ancien Régime, le théâtre parisien reste un lieu public bigarré et bruyant, mais il se discipline, les armes y sont interdites aux laquais, les chiens expulsés, les siffleurs patentés mis à la porte, enfin les spectateurs du parterre sont assis, et ainsi plus *difficiles à mouvoir*.

C'est l'accroissement de ce public qui fait la fièvre théâtrale de Paris au XVIII$^{e}$ siècle: Louis Lagrave évalue à 400 000 ou 500 000 le nombre des entrées annuelles vers 1750. Il justifie le retour des Italiens, redevenus, en 1723, *comédiens ordinaires du Roi*, parmi lesquels triomphent Luigi Ricoboni et Marivaux, il sous-tend la montée décisive des théâtres, des foires et des boulevards, où vont s'opérer une autre répartition des publics et s'élaborer des expériences dramatiques nouvelles. La hiérarchie des théâtres, imposée par la tradition et par l'autorité, correspond en gros à celle de la société, l'*Opéra*, la *Comédie* sont moins bourgeois que les *Italiens*, il y a une distance entre les Halles et le faubourg. Paris est fou de théâtre, et la moyenne des affluences, calculée pour tous les spectacles, ne fait que croître entre le règne de Louis XIV et celui de Louis XVI. De l'affaire, les architectes songent à multiplier les projets et à imaginer de nouveaux dispositifs qui inspirent les modèles italiens, puis les modèles antiques. L'incendie de l'Opéra en 1763 et 1781, la vétusté de la vieille Comédie française suscitent "les *théâtres de papier*" de nombreux architectes, puis les initiatives intéressantes quand la ville et la maison du Roi s'entendront pour laisser créer de nouvelles salles, à l'instar de la province. La première salle nouvelle revient à l'initiative privée, c'est le théâtre construit en 1752 par Monnet pour l'Opéra comique de la Foire Saint-Laurent. Après 1770 sont reconstruits, ou construits, l'Opéra, la nouvelle salle de la *Comédie française*, à l'Odéon, dans le façonnement d'un quartier expressif du nouvel urbanisme (D. Rabreau), puis la nouvelle comédie italienne vers 1782; le théâtre est alors devenu monument de l'urbanité.

## Apollon en province

En province, le cheminement suivi avec quelque retard est celui de Paris, mais l'entregent des élites provinciales leur confère au XVIIIe siècle un rôle novateur spécifique. Au départ, il y a peu de théâtres implantés en permanence et aucune salle affectée exclusivement au jeu des troupes de passage. Les comédiens utilisent leurs ressources locales qu'autorisent les autorités, toujours maîtresses d'accorder ou de refuser la liberté de jouer. A la fin du XVIIe siècle apparaissent les premiers théâtres construits pour l'Opéra et ses machineries, après les succès parisiens de Lully et de ses successeurs: Rouen en 1682, Lille en 1702, Strasbourg en 1701-1702, Metz en 1703, se dotent des nouveaux équipements. L'itinéraire des troupes suit plutôt le réseau des protections et l'appel des amateurs citadins dont la fidélité entraîne les passages répétés des mêmes comédiens. En dépit de l'aventure fameuse de l'*Illustre Théâtre*, qui visite entre 1655 et 1657 Pézenas, Narbonne, Bordeaux, Agen, Lyon, la moitié nord de la France a été plus fréquentée que la moitié méridionale. Les troupes passent 72 fois à Dijon, 57 fois à Lyon, 53 fois à Rouen, à Nantes, 48 fois à Lille; au sud de la Loire, 22 fois à Marseille, 21 à Bordeaux, 11 spectacles seulement à Toulouse. Trois raisons rendent compte de ce décalage: la trame urbaine et l'état du réseau routier, le progrès de l'alphabétisation et la constitution des élites culturelles, la géographie des patois et de l'extension du bilinguisme. Cependant, la vie théâtrale méridionale connaît sans conteste son apogée avant 1650 avec un répertoire en langue d'oc, un public composite, rassemblant les bourgeois et le populaire, des auteurs de talent tels Bruey, Zerbin et Fébau, des acteurs renommés qui itinéraient du jeu de paume en salle d'auberge, et souvent, un succès local centré autour des petites capitales (E. Leroy Ladurie): Aix avec Bruey avant 1650; Béziers avec le théâtre des *caritats de l'ascension*, au temps de Louis XIII; Avignon jusqu'à la Fronde; Agen, Montélimar, Cahors, Montpellier, Carcassonne, Nîmes, Marseille, qui éclipse les autres cités; jusqu'à Grenoble et même Limoges. La carte du XVIIe siècle enregistre les étapes d'une conquête du royaume par un répertoire, une langue, un style de jeu: le théâtre parisien en français qu'interprète les troupes libres et protégées, celles des Conti, des Villeroy, des Condés, du Duc d'Epernon, de Monsieur, qui parcourent le sud et le sud-ouest du royaume. La carte du XVIIIe siècle prouve le ralliement des élites urbaines dans le Midi comme dans le Nord. Le répertoire occitanien se maintient quelque peu mais dans un registre déclassé, qui bénéficie de l'unanimité festive passagère des notables et du peuple: celui de la tragi-comédie baroque où s'expriment l'amour de la vie et les sentiments ordinaires. Pour le reste, l'Occitanie urbaine privilégie aussi un théâtre en français. C'est le temps où se fixent les troupes et où se construisent les théâtres.

Dans le réseau urbain, une dizaine de cités bénéficient d'une activité

théâtrale continue, depuis la fin du règne de Louis XIV et en dépit des difficultés des temps: Lille, Amiens, Bayonne, Nancy, Avignon, Bordeaux, Nantes, Besançon, Dijon et Rouen; une autre douzaine de cités reçoit des troupes par intermittence. La province se dote de salles de spectacles modernes et dans certains cas exemplaires par rapport à Paris. Le théâtre quitte les jeux de paume, les hangars, les granges et les salles de banquet pour s'installer dans ses meubles, avec commodité et quelquefois luxe. En 1789, D. Rabreau évalue à 50 le nombre des villes qui se sont dotées d'une salle fixe dans un ancien local, d'une nouvelle salle et d'un grand théâtre *temple des Muses* (huit pour ce modèle à Besançon, Bordeaux, Dijon, Lille, Marseille, Nantes, Nîmes et Strasbourg après 1780). Pour les quatre cinquièmes, ce sont des villes de plus de 10 000 habitants, où l'on retrouve tous les facteurs du développement de la vie culturelle (cours souveraines, collèges), mais aussi activités négociantes et garnisons, car le théâtre est porté à la fois pour le goût de l'ancienne noblesse et celui des nouvelles élites. Les troupes sont partout décentes et le répertoire est celui de Paris; on joue, en alternance, tragédies, comédies, opéras, opéras comiques, pantomimes. Derrière le plaisir et le loisir recherchés par les notables urbains qui encouragent les fondations et les finances, trois enjeux se profilent qui traduisent l'évolution de la société (D. Rabreau): la police urbaine, pour un meilleur usage et un meilleur contrôle des lieux de spectacle, la spéculation, le prestige culturel.

Il s'agit d'abord de discipliner le public, tout en le distrayant, par la surveillance policière ou militaire, par les règlements de *police*; l'effervescence des spectateurs est canalisée, on empêche le détournement des lieux vers d'autres activités, le jeu, la prostitution, la boisson, ainsi l'on vite le débordement des cabales conduites par les différents composants sociaux du public: nobles contre bourgeois, populaire contre élite. La spéculation immobilière et le commerce installent le théâtre dans un édifice public: il devient une affaire entre commanditaires, actionnaires et entrepreneurs. Le théâtre provincial, en effet, n'est pas protégé: c'est un secteur de liberté et de concurrence. De surcroît, il permet maints profits, car souvent sa construction entraîne des aménagements ou l'extension vers des terrains voisins de la spéculation; c'est le cas à Lyon avec Soufflet, à Bordeaux, à Nantes, à Marseille, voire même au Havre et surtout à Paris, autour de l'Odéon. Ni le bon plaisir du prince ni le fonctionnalisme théâtral pur ne guident désormais les constructions, mais une économie de *progrès*, liée à l'utilité publique; une politique urbaine où en marge de la société des ordres agissent des citoyens qui s'associent par libre contrat pour un profit et un loisir. Enfin, la moisson culturelle s'engrange dans les nouveaux temples de l'art: la gloire du prince s'y déploie dans le répertoire parisien et l'éducation des peuples pour un ordre policé. Les philosophes y ajoutent la morale civique et le patriotisme. L'inauguration d'un théâtre comme la séance publique d'une académie est une cérémonie de liturgie civique.

C'est un autre élargissement des pratiques culturelles des milieux notables et l'affirmation symbolique de la cohérence d'une société par-delà ses divisions.

## De la foire au boulevard

C'est que la théâtromanie est partout, elle s'exprime dans les spectacles de la cour comme dans les théâtres privés, elle mobilise les riches, mais aussi le peuple, elle fait les auteurs à succès. Rousseau prêche dans le désert, car jamais société n'a été aussi enjouée de théâtre que celle des Lumières. C'est l'expression d'un besoin qui traverse toutes les couches sociales, c'est un enjeu politique, car par la généralisation de la fréquentation du théâtre se prépare le règne du public et d'une opinion, en attendant les vraies libertés. C'est pourquoi le destin du théâtre de foire mérite attention. A Paris principalement, en province aussi, à la foire de Guibray ou à celle de Beaucaire, à celle de Reims qu'a étudiée René Gandilhun, il est lieu de rassemblement et de contact, il active le commerce des choses et celui des hommes. Le théâtre prend rang dans une pédagogie de la consommation qui s'universalise. Durant tout le XVIIe siècle se sont maintenus dans ces assemblées périodiques les gestes de la culture de la place publique. A Lille, Chavatte est au XVIIe siècle le chroniqueur précis des exhibitions de bêtes fauves ou d'animaux savants, des jeux des nains et des géants, des prouesses des *banquistes*, des mangeurs de feu. A Paris, la Foire Saint-Germain et la Foire Saint-Laurent voient se succéder les spectacles des jongleurs, des sauteurs, des marionnettistes, et la multiplication des troupes foraines venues ici se remplumer après des tournées provinciales misérables et faméliques. L'expulsion des Italiens donne un coup de fouet aux forains, les entreprises se multiplient, avec les frères Allard, Domenico, la Veuve Maurice, Bertrand, Gauthier de Saint-Edme et Octave. C'est aussi un secteur de liberté qui obtient un franc succès, le public en est friand et les spéculations, bourgeois de Paris, financiers comme Nicolas et Baron, les paumiers toujours intéressés par les représentations, investissent dans les entreprises foraines. Celles-ci, en butte à l'hostilité des comédiens français qu'appuie la justice, s'adaptent: ils doivent se contenter de monologue puisque tout dialogue est interdit, puis du jeu à la muette, où les acteurs miment un texte écrit sur les écriteaux que déchiffrent les lettrés du public. Le théâtre de la foire traduit dans ses formes dramatiques l'écart social et culturel qui sépare la culture du petit nombre et celle de la Place publique. Dans la réalité, il permet aux pratiques originales du théâtre forain de jouer un rôle intermédiaire entre les domaines culturels. Au terme de bien des avatars, c'est la liberté qui l'emporte. L'union de la tradition italienne avec celle de l'opéra comique

forain, qu'avait soutenue Monnet, se concrétise après 1762, marquant le succès de l'imagination sur le monopole. Jusqu'à la Révolution, le théâtre des boulevards en multiplie les effets, avec Nicolet, Gaudon, les Grands danseurs, les Variétés, l'Ambigu comique. C'est la vogue prérévolutionnaire des *spectacles inférieurs*, qui ont relayé les tréteaux forains. Pour ce théâtre ont travaillé des auteurs de talent, oubliés parce qu'a triomphé dans la culture le *grand théâtre*: Lesage, Fuzelier, d'Orneval, Piron, Favart, et quantité d'hommes de lettres du second rayon, composant pour des entrepreneurs rapaces, Audinot, Pompigny, Maillé de la Malle, Saint-Aubin, Mayer de Saint-Paul: ce sont les pauvres honteux du Parnasse et les Rousseau du ruisseau, champions d'un théâtre mineur.

Le public est ici populaire mais pas uniquement: tout le monde va à la foire et au boulevard, la bonne compagnie elle-même tend à s'imposer aux spectateurs de plein vent et de parades. Monnet sollicite pour l'opéra comique une ordonnance de police qui codifie la décence de son spectacle et de sa salle à l'instar de celle des autres grands théâtres. L'important reste que la bonne société n'a pas réussi à confisquer totalement le succès de la foire et du boulevard. Là se joue le goût des gens de peu pour un style dramatique appuyé et la fascination des honnêtes gens pour les formes d'expression populaire. On s'encanaille devant la parade de Nicolet ou dans la salle de Monnet. Ménétra, pour qui le théâtre joue un rôle essentiel, y côtoie ceux de la cour et de la ville. C'est pour les autorités un espace de déclassement moral; c'est l'effet *mylord l'Arsouille*, qui joue pleinement pour plusieurs raisons. L'esthétique foraine en effet conforte l'union des effets qui sont séparés dans la hiérarchie des grands théâtres, c'est aussi un spectacle qui met en jeu tous les sens, c'est enfin un théâtre du merveilleux. En même temps, cette esthétique hybride utilise des procédés qui frondent les grands genres: c'est un théâtre de la force et de la souplesse, qu'incarne l'ambigu "*Arlequin*", c'est un théâtre du sous-entendu et de la connivence, royaume de la scatologie et de l'obscène, plus ou moins appuyés et voilés, c'est le théâtre du détournement et des ruses qui suppose, au temps de la persécution par la Comédie française, une participation unanimiste, avec lecture, change et collaboration des lettrés et des analphabètes. Selon la place qu'on occupe dans le théâtre et dans la société, le répertoire forain et son jeu peuvent suggérer la critique, la fronde, ou garantir l'ordre et la stabilité en réduisant le message contestataire à un effet de distance ou de stéréotype. D'où, sans doute, l'importance de la *Parodie*, qui fait rire tout le monde mais divertit sérieusement le petit nombre des amateurs, qui iront jusqu'à écrire eux-mêmes des parades pseudo-populaires, pour leurs jeux particuliers; Ainsi le spectacle forain doit son succès au mouvement qui uniformise la culture de la société classique, il fait descendre vers le bas une part de la grande dramaturgie, mais aussi, remonter vers les classes dirigeantes les pratiques culturelles de la place. Dans une proxémie neuve, l'érudition et l'ironie élargissent une

distance culturelle. Jean-Jacques Rousseau ne s'y trompe pas, en opposant les théâtres aux fêtes de *fraternité publique*, la tristesse à la joie, les ombres du théâtre à la réalité de la vie.

## Bibliographie

H. C. Lancaster *A History of French dramatic Literature in the XVII century*, 9 vol., Baltimore, Johns Hopkins, 1929-42.
H. C. Lancaster *The Comédie française, 1680-1701*, 2 vol. , Baltimore, Johns Hopkins, 1941.
J. Lough *Paris theatres audiences in the 17th and 18th centuries*, London, Oxford University Press, 1957.
P. Mélèse *Le théâtre et le public à Paris sous Louis XIV*, Genève, Droz, 1935.
G. Mongrédien *La vie quotidienne des comédiens au temps de Molière*, Paris, Hachette, 1966.
*Histoire de la France urbaine*, tome II, Paris, Seuil, 1981, réd. par G. Duby et J. Le Goff.

# Charity and Family Subsistence: Florence in the Early Nineteenth Century

STUART J. WOOLF
Essex/Florence

Charity implies choice; and choice, in turn, usually requires some method of checking. If we leave aside casual or indiscriminate distribution of alms (say, of small change to beggars outside a church, or soup to the needy at the convent gate), the regular and above all the continuous exercise of charity has always necessitated discrimination because of the permanent disproportion between resources and demand. In this sense, charitable institutions in the past functioned in similar manner to contemporary departments of social security, elaborating mechanisms to identify preferences: personal and family information was demanded and classified, such information was checked against anticipated norms or even by direct investigation, and on this basis either rejection or an appropriate level and form of assistance was decided. The distinction habitually made between charity and the welfare system is that, for the latter, need is the sole criterion giving entitlement to assistance (at least until the attacks on the Welfare State of the past few years), whereas for the charitable organizations of early modern and modern Europe value judgments of worthiness accompanied assessments of need. In fact, as the literature on the persistence of poverty in the Welfare State has amply demonstrated,[1] assessments of need are often difficult to arrive at, even with the imposing statistical back-up of modern government, and easily contain value judgments. What is common to contemporary social security organizations and earlier outdoor relief charities is that because both work from the premiss of family needs, they are obliged to elaborate models, or at least Weberian ideal-types, of different family structures or different phases in the family cycle, in relation to their relative earning capacities, against which to test needs and upon which to draw up their regulations. It is this identification by an early nineteenth-century Florentine charity of family structures, cycle and subsistence needs that is the main concern of this paper.

The Congregation of San Giovanni Battista was the main outdoor relief organization of Florence, to which a third or more of the poor turned for assistance – an annual average of 8 600 in 1810-12, or over 10% of the urban population.[2] By statute, it only assisted the labouring poor of the

city, excluding not only non-Florentines, but citizens unable to prove their domicile, some job-skill (however rudimentary) and moral conduct. It was also statutorily limited in the outdoor relief it could offer, which consisted of five objects – cash subsidies, beds, bedclothing (sheets, occasionally blankets), clothes and bread – some of which were subject to further conditions. These limitations derived from the highly specialized organization of charity in Florence (as in most Italian and western European cities), by which individual institutions assumed specific responsibilities (such as hospitals for foundlings, conservatories for orphan girls, etc.). But our Congregation played a central role in Florence of the eighteenth and nineteenth centuries, as it existed to support the fundamental unit of social life – the family in its home – at moments of particular crisis. Permeating all its assistential activities was the concern to underpin the independence of the family as a unit, to prevent individuals or entire families from becoming permanently dependent on institutional support (for instance, in hospitals, workhouses or asylums), or alternatively from adopting strategies of survival disapproved of by society (theft, prostitution, unauthorized begging, etc.).

In an earlier study of poverty, family structure and charity at Florence, I suggested, on the basis of a statistical study, that the Congregation concentrated its efforts on critical moments of the family cycle.[3] The purpose of the present paper is to explore this conclusion on the basis of a non-statistical analysis of the descriptions and comments made by the deputies of the Congregation about each application for assistance. As in my previous article, the evidence is based on a sample consisting of all the applications from a single *sestiere*, numbering 1 219 family units (4 498 individuals) or 23% of the series. This second *sestiere* covered the *quartiere* of Santa Maria Novella and part of the zone around the cathedral of San Giovanni, typically popular areas of Florence, densely inhabited by artisan and street-trading families.

In order to request assistance, the Congregation required applicants to fill in a printed form, listing information about each individual living in the household: name, relationship to the head of household, age, state of health, job-skill, employer and weekly wage (if any); underneath these details, the applicants specified the form of assistance they hoped for. At the back of the form, the parish priest was asked to confirm the information and add his own comments, which were followed by those of two deputies of the Congregation. It is worth dwelling briefly on these procedures of acquiring information, as it was solely on their reliability that the individuals, units or categories to be assisted could be identified. Four aspects can be noted.

First, the quantity of information demanded was considerable and can be compared, without impropriety, to that required by contemporary social service offices. In a substantially illiterate society, the compilation of

these forms must have frequently (usually?) required the collaboration of someone outside the household (parish priest? deputy of the Congregation? a neighbour or relative?), whose very participation can be read by us ambivalently as implying either a check on the veracity of the information or collusion with the supplicant (or both). But, perhaps more importantly, the quantity of information required must have discouraged the casual applicant (that Florentine yesteryear equivalent of the "scroungers" living off social security, so useful to our English and American neo-liberal critics of the Welfare State). It could have been no light matter applying to the Congregation, not something to be tossed off, like a football pools coupon, on the off-chance of a useful supplement to one's income; but a desperate measure, often a final resort, in the face of total indigence, destitution or worse. Indeed, the humane, upright and conscientious deputies of the Congregation must have reinforced this austere impression by their automatic dismissal of applications "because of inadequate information", or its "improper presentation", not only by the supplicants, but by the parish priest ("because of the limited information provided by the priest", "because the curate has written his usual general phrase"). Full presentation of this substantial corpus of information was thus a bureaucratic prerequisite, failure to comply with which risked the chilling response of one deputy: "Refused in that providing little information implies no great need".[4] Only the genuinely needy or skilled rogues (our present-day "scroungers") were likely to go the trouble of applying. And the latter, although they did not know it, were rapidly identified in the deputies' registers and rejected, if they had ever before obtained assistance ("already assisted other times").[5]

The second point to be noted about the applications is the character of the data demanded. The information about family relationships and age (common to many census-type sources) was essential to the deputies, in a general sense to set against their ideal-type families, and in an immediate sense to judge the nature and urgency of their intervention (e.g. separation of pubescent daughters from the parental bed). State of health and job-skills were more unusual, both serving (like weekly earnings) as component elements towards a composite judgment about levels of vulnerability; but also useful as discriminatory information, both negative and positive: to decline responsibility for those without any job-skills (another statutory prerequisite), to elicit information about alternative supplementary benefits (was the applicant in receipt of "the usual subsidy for new-born babes from the Innocenti hospital"?), to recognize the drastic consequences of the illness or even worse the hospitalization of a head of household ("the grave and serious illness of the head of the family").[6] Earnings, both individual and of the household unit, obviously represented a crucial element in the deputies' decision and (as we shall see) were subjected to some scrutiny.

The third aspect we must note is the procedure of visitations. A few

families were at once excluded because they lacked the parish priest's confirmation of good conduct: "The family is not of good morals and for this reason has not been visited".[7] On receipt of an application, two of the 72 deputies proceeded to a visit of the room in which the applicants lived. Such personal contact was an essential ingredient of nineteenth-century Catholic charity, shortly to be fixed in its definitive mould by de Gérando in his best-selling *Le Visiteur du Pauvre* (1820) as the most effective method "of recognizing true indigence, and rendering alms useful to both donors and recipients".[8] In practice, the visitation permitted the deputies to assess the applicants' living conditions (although presumably not to verify earnings). A negative judgment from the deputies ("very poor, but not of the most needy") was virtually equivalent to rejection.[9] Following the comments of the visitor-deputies, another deputy, acting as rapporteur, wrote his proposal for a decision, which was then submitted to the Congregation's bureau (six deputies, acting in rotation). This rapporteur's proposal was almost always clearcut, accompanied by a brief explanation. But on rare occasions he expressed uncertainty and left the decision to the bureau, as in the case of an unauthorized beggar (technically disqualified from further assistance) whose family lived "in great and extreme misery", or the wife, forced by her husband's hospitalization to change to a cheaper but totally empty room, asking for a bed and bedclothing which the regulations did not permit.[10] Applications were occasionally rejected in the absence of adequate information from the visitor-deputies ("given the little verification of the request by the deputies"), even through their failure to check the age of a starving widow.[11] An obligatory requirement for a subsidy to pay for a wetnurse was a visit by a doctor, sometimes (always?) the Congregation's doctor, whose report occasionally led to refusal.[12]

This complex procedure based on personal knowledge brings us to the final point – verification of the veracity of the information. Here the most delicate item was the declaration of earnings. There were no direct means by which the deputies could check the amounts earned from intermittent (mostly textile) putting-out work or street peddling; nor, in this instance, could they rely on the parish priest, little better able than they to check and more likely to confirm than to deny (and hence ensure the rejection of) the declarations of worthy and genuinely needy parishioners. The deputies, in fact, were suspicious and rejected a fair number of applications on the explicit grounds of "greater earnings than those declared", "must earn more than declared", "untrue declaration of personal earnings", "false declaration of earnings", and similar verdicts. Usually we do not know on what evidence they based their conclusions. But on occasion they point to specific skills, such as the silk-weaver's (where work was not lacking and hence earnings must be higher than declared) or the coach-builder's (who must be earning something); or deny that it is

possible for a family to declare no earnings; or comment that the age and good health of the individuals mean that they must be earning more than they declare.[13] There can be no doubt about the scepticism of the deputies over the precise earnings declared, nor about their direct knowledge of wage levels. But not only is the proportion of applications rejected on these grounds minimal (perhaps 60 out of the total 1 219); more important is the statistical evidence that the declared earnings of three quarters of all applicant families provided less than minimal subsistence, so that even substantial under-declarations would not have removed them from the most vulnerable category of poor.[14] At best, gross or inept under-declaration of earnings served the deputies as a further means of discrimination amidst a multitude of desperately needy families.

That the deputies were aware that virtually all applicants were not only poor but in critical circumstances is evident from the terminology they employed. "Poor" (*povero*) described a permanent condition, "wretched" (*miserabile*) a state of urgent need, a "poor" family was presumed able to find the means of ensuring its subsistence, a "wretched" one was clearly beneath the poverty line. For the families themselves the distinction between *povero* and *miserabile* was tenuous and mobile, a continuum rather than a contrast, for both labels not only included families practising the same (usually low-level) skills and living in the same houses, but could describe the same family at different moments of its cycle. For the Congregation, the distinction was definitive: "Poor to the utmost, but not of the most needy. Rejected".[15] The operative words for serious consideration by the applicants – were "wretched" or "needy" (*bisognoso*), around which was woven a miniature tapestry of qualifications: "wretched enough", "very wretched", "extremely wretched", "most wretched", "state of wretchedness", "truly needy", "state of extreme need", in short, appurtenance to what was significantly called "the class of the wretched". Given these variations on the term *miserabile*, visitor-deputies were obliged to employ alternative signals to highlight the most urgent cases, such as "maximum indigence" or "total destitution".[16]

It is evident that such terminological distinctions served a function of discriminatory choice, in exactly the same way as the rejections through inadequate information. The deputies never explain what they mean by "poor", at most affirming that the applicants earn or, given their condition, could earn enough for their subsistence. It is rare to find so explicit a statement as "not appropriate to be assisted, as sufficiently provided with everything".[17] In like manner, and far more frequently, the deputies took refuge behind their regulations. A substantial number of rejections (38% of all applications were rejected) were justified generically as "contrary to the regulations". In specific instances the refusal might be because the application was for something not included among the Congregation's items for outdoor relief – a dowry for a daughter, assistance to train a son

as coach-driver. Alternatively, the application would be rejected because it came from a beggar, whose licence to beg from the Congregation automatically precluded further assistance. Families applying for blankets in the summer were not granted them as it was not yet the right season.[18] Yet again, the rejection might be based on the non-conformity of family conditions with the specifications of the items requested – the fact that a mother was actually suckling her new-born babe, even though she was starving, when the regulations stipulated a physical inability to breast-feed; clothing for adults, even when they were literally in rags, as only children were statutorily entitled; refusal of a bed, because the children were too small to require separation, even if there were many of them.[19] As one would expect, some rapporteurs were more meticulous than others in their interpretation of the rules. It is possible to point to extreme cases of bureaucratic-minded deputies who had lost sight of the human distress staring out at them from the applications, such as the refusal of a wetnurse subsidy because the babe's incapacity to suckle (as distinct from the mother's to feed) was not foreseen in the regulations. Red tape most clearly smothered the direct contact on which the charity operated on those rare occasions when the rapporteur decided primarily or exclusively in terms of maintaining a rough balance of charity between the parishes: "refused, given the multiplicity of subsidies needed by this parish [San Lorenzo]", "granted, given the few subsidies that this parish [Ognissanti] has obtained".[20]

But it would be misleading to place too much weight on a legalistic approach to the requests of these needy families. The deputies were not only compassionate, but were prepared to interpret the rules broadly, even in special cases to go against them. Mothers denied the wetnurse subsidy were often granted nappies instead. Single old women were frequently given at least bread. Very large families, with four, five, up to nine children, were granted beds even if the children were not of the statutory age. But even small families could reasonably hope for a bed, if their living conditions were particularly shocking ("they are all sleeping together on a straw mattress on the damp floor", "sleeping on boards", or "on a chest", "on straw", "more than a bed, it seems like a dog's kennel", "on the bare floor", etc.).[21] Grants were occasionally made despite inadequate declarations, "because of the information provided by the parish priest". A wetnurse subsidy was granted, although not within the regulations, "given the wretchedness of the family and the lack of employment"; a family "wretched to the utmost", denied a bed, was given an alternative subsidy. A bed was granted, even though it was noted that the father was an authorized beggar; or where father and son were sleeping on the bare floor, "even though it is known that this family already got a bed from the Congregation, which perhaps was sold". A wife about to give birth was often sufficient justification for the grant of a bed. Most revealing of a

willingness to interpret the rules in order to fit the human case are two examples of beds which should have been refused as the daughter was marrying. In the first case, it was noted that the bed was the property of the daughter, whose marriage would have left her father and sister bedless; in the second instance, having observed that the daughter was about to marry and could not do so as she was the only earner in the family (and hence unable to buy a bed), it was decided, casuistically, that a bed could properly be granted to separate her from her parents and not because she was marrying.[22] If the regulations were sometimes used as a means of negatively determining choice, if individual rapporteurs could sometimes display deeply hostile suspicion and reject applications perhaps in a bout of irritability, there is abundant evidence that for the most part the deputies utilized the evidence they had assembled so carefully in order to assess the relative degrees and urgency of need of the families they had visited.

What did the deputies look for? What criteria did they employ to guide their choice? The comments they wrote on the applications provide us with a fairly definitive picture which – bearing inmind that the information was based on face-to-face contact – can reasonably be accepted (at least forthose aspects which they noted) as an acurate description offamily and living conditions in these lower, more unfortunate levels of the poor.

Three parameters can be identified, within which the deputies located their entire discourse – the presence of kin, the level of earnings and the size of family. All three, separately or in various combinations, were repeatedly employed as a means of explaining *why* a family was *miserabile* (and hence meritorious), or why it should not receive help. The washerwoman, living alone, was denied help as her son, even if not living with her, could provide for her subsistence; the aged widowed mother, with low earnings, merited assistance as her only son was about to leave her in order to marry; the old, mentally ill widow, living in the same house as her married daughter, was granted aid as her son-in-law (himself an applicant for a bed) was not able to help her. A conscripted son was a guarantee for a grant, precisely because conscription implied enforced absence of kin (even worse, of male kin) and hence a sometimes drastic diminution of earnings.[23] The quintessential prerequisite for an application was inadequate earnings, repeated in virtually all applications, whose counterproof is to be found in the systematic refusal to grant aid because in the deputies' judgment the family earned enough. Size of family functioned in similar dualistic manner: "maximum indigence" was the alarm-bell sounded for big families (nine, even eleven members) "because of their large number and low earnings"; whereas the single individual, even female, was refused aid "because she is healthy and of an age to earn her living".[24] Within these three parameters, state of health provided a crucial supportive function, rarely sufficient of its own to justify assistance, but a clinching argument in combination with one of the three permanent conditions.

Thus the ageing man of limited health was rejected because he had no family to support; the family of three headed by a sick father was assisted because of inadequate earnings; husband and wife were given a grant, as the husband's earnings, normally adequate, were insufficient to pay for all the medicines his wife needed during her prolonged illness.[25] Two ideal-type meritorious families, at the extremes of our continuum, were the single widow, alone in the world, aged, of uncertain health and hence unable to earn enough for subsistence; and the large, sometimes extended family, with children too young to earn, an aged parent, and one of the normal earners sick.[26]

A certain symmetry thus existed between family structures likely to provoke cool detachment and those certain to attract sympathetic attention among the deputies. A check-list of the former, destined to be refused assistance, would read as follows: "excellent earnings" or "sufficient earnings"; "young, healthy and [hence] capable of earning"; a young couple without children; an adolescent of sufficient age to earn enough; a single young woman; an old but healthy man or widow; a female head of household aided by her children; a female with small children requesting a bed, or with a son old enough to buy one; a daughter leaving home to marry asking for a bed. Certain job-skills, judged as generating adequate income, were likely to justify rejection of those who declared them: female silk-weaving especially, but also washerwomen and an innkeeper; and servants, as they were expected to be looked after by the household that employed them.[27]

The counterpart to these types of families regarded by the deputies as self-sufficient was the family condition that merited particular attention. Single aged women, if ill, were almost certain to receive at least bread (if healthy, they might be rejected on the grounds that they could beg); orphans were guaranteed help, as were abandoned wives; the recent death of a husband could ensure assistance, as would the additional burden of an aged infirm parent. Children too young to earn, a deranged member of the family, serious or prolonged illness of a resident relative, and especially of the head of household, all guaranteed assistance; large families were almost equally sure of receiving a bed, not least because they were likely to include at least one pubescent daughter who had to be separated from her male relatives.[28]

It would be over-simplistic to conclude that the deputies of the Congregation restricted their help exclusively to the critical moments of the life and family cycle. Social and especially moral considerations played a not inconsiderable role. Families with some link with the Congregation or the parish priest were especially meritorious; recommendation from an influential person (the local police official, a high-level bureaucrat) could prove decisive. Soldiers and officers, suddenly unemployed after 25-30 years' service through the vagaries of political change, were given sometimes

substantial help. Professional skills – as a copper engraver, a violinist, a teacher – deserved particular regard, as did the dismissed clerks of the former grand-ducal administration. Occasionally and imperceptibly such social considerations could merge into traditional and by now anachronistic concerns for social status, for the shamefaced poor (even though the legal responsibility for the shamefaced belonged to the separate institution of the Buonomini di San Martino). It is difficult to tell whether a dismissed clerk was genuinely destitute or merely unable to sustain his proper station. But there can be no doubt about the families described as of "good birth" or "civil condition", regularly granted assistance; nor about the family "of noble birth", where the wife fell ill through having to do all the housework, so obliging the husband to pawn his possessions in order to pay for a woman to look after wife, three children and home.[29]

Once more, we must not exaggerate this handful of instances of social considerations reinforcing, and in rare cases even replacing, assessments of the economic condition of the families. Certainly more important were moral considerations. In extreme cases, usually of large families, an unruly child or one who refused to work would be removed from home and placed in the disciplinary Casa Pia.[30] Daughters were given clothes in order to go to church.[31] But above all beds were granted to separate the sexes, even when it meant going against the regulations or despite fears of fraud. Daughters entering puberty were the main concern, but sons reaching an age of sexuality (in one case noted as 14) also provided an impelling reason to grant a bed. The deputies were quite explicit: the daughters were of "a dangerous age", it was "a scandalous matter" for daughters to sleep together with young parents; on the contrary, a six-year-old girl could continue "for some time" to sleep with her parents and maternal grandfather.[32] Potential sexual promiscuity was the more worrying as kin sleeping together extended beyond parents and children: orphan nieces needed to be separated from their uncle, a father-in-law and his son from the married couple, children and grandfather from the marital bed – while a pubescent niece was told to continue sleeping in the same bed as her uncle, given his old age and infirmity.[32]

Specific moral considerations thus coloured the deputies' judgments, though even in the case of sexual worries they did not often lose sight of the possibility for the family to buy its own second bed. Nevertheless, their prime concern remained the economic independence of the family throughout its cycle. This explains why they never refused to redeem the household possessions pawned by a family at the *monte di pietà*, as their sale by the *monte* marked a significant rent in the thin fabric of the family's autonomy.[34] The deputies were always sensitive towards negative conjunctures in the family cycle: illness, unemployment, an imminent childbirth, loss of a son through conscription or a husband through death were regularly accepted as justification for granting assistance. But, more

importantly, they also displayed a clear awareness of the structural relationships between levels of earnings and, on the one hand, the individual life cycle (in which age and gender were the determinants); and, on the other hand, the family cycle (in which income was limited not only by age and gender, but by the number of children and the small proportion of earning members).[35] Large families and single old women were given a preference precisely because they were the least likely to earn enough for subsistence. But the most miserable job-skills, such as muck-raking or street peddling, were recognized as yielding "uncertain earnings", or as unlikely to provide enough for a family.[36] Unequivocally the policy of the deputies consisted of intervention to underpin the most vulnerable family structures and the most critical moments of the family cycle, which they identified by direct impressions and experience, without need of the statistical evidence of subsequent historical research.

In a simple model of urban poverty, charity would thus have been channelled to discrete nuclear families, each comprising its own household. In my earlier article I pointed to the practical identification of family and household: 94.6% of the applications described two-generational nuclear families, and a mere 0.6% included non-relatives or more distant kin.[37] Analysis of the deputies' comments, however, reveals a more complicated image, as a small minority of applications came from, or on behalf of, individuals living (and sleeping) with larger families. Although the identification of household and family remained overwhelmingly the norm, it was the abnormal cases that attracted disproportionate attention from the deputies, precisely because of the burden represented by the additional household members.

Among this limited number, the most frequent case is that of an aged parent: the mother living with her son's family of four, unable to support her, merited help, as did the son-in-law, unable to support his mother-in-law when she fell ill, even though she did not live in the same house.[38] The principle underlying these favourable decisions was that acceptance of family responsibilities could overstretch subsistence capacities; we can consider it as a variant on the assistance given to widows living alone, who did not deny that they had kin, but described them as in as wretched a state as themselves. Another variant was when the family was so poor that a newly married son or daughter could not afford to rent a separate room: bedding complications were created, in which not just sisters, brothers, mothers or fathers, but even fathers-in-law and brothers-in-law found themselves sharing the same bed, and appealed to the willing ear of the deputies.[39]

More complicated were the cases where blood-ties were less close or even non-existent. Orphans were the main group concerned: the niece sleeping with her married aunt, with husband and four children, was an obvious candidate for a bed.[40] But if the orphan was without relatives, the

importance of finding a family ready to accept her (for almost only cases of girls are recorded) more than justified a bed.[41] There could be perfect consonance between the ideals of the Congregation and the applicant family, as with the widow Mazzaranghi, who had taken in the Braccini orphans, brother and sister, and "it seems, is giving them an excellent education".[42] But in other instances, the orphan might only be taken in because of the householder's need of assistance: widow Pestellini, old and ill, was obliged to take in a girl from the public orphanage (Bigallo) to share her work (and hence her insignificant earnings). The orphan Anna Chini, "sick and infirm", lived with non-relatives, but in such wretched conditions that she was granted aid, "given her pitiful situation".[43] The examples of households containing more than the classical nuclear family are too few and the comments of the deputies too cryptic to deduce any elaborate significance about the implications of cohabitation at this level of subsistence, where space, beds and food were all in very short supply.[44] But it is evident that the Congregation's prime concern was to find a home for orphans and that it was prepared to listen sympathetically, even to stretch its rules in order to achieve this end.

What conclusions can be drawn from this study of charity and subsistence? There can be no doubt that the Congregation of San Giovanni Battista understood its duties of outdoor relief (*a domicilio*) in a very literal sense: the deputies saw their task as underpinning the independence of the family, both economic and residential, at moments of crisis. To achieve this end they were prepared to intervene at various levels of subsistence crisis (from a family's initial use of the pawnshop to its sudden awareness, through absolute lack of food, of the immediacy of starvation[45]), at different stages of the family cycle (orphans, large families, aged widows, etc.), and at particular conjunctural moments (sickness, insanity, unemployment, etc.), while deliberately excluding others (young couples, healthy adults, individuals with skills judged capable of ensuring adequate earnings, etc.). Precisely because of their overriding concern with the family in its home, they were particularly sensitive towards the problem of orphans, who lacked kin and consequentially also neighbourhood support. But equally, although they preferred to deal with entire families (and hence insisted on full information), they were prepared, wherever judged necessary, to intervene within households in order to assist a specific individual. The conviction of the deputies about the centrality of the social role of the economically independent family unit would be difficulet to fault. At most, one may query whether the forms of charity they proposed were appropriate to the world of structural poverty which was their domain.[46]

Institutional sources offer little direct evidence about how families viewed and responded to the cyclical subsistence. In this instance, application to the Congregation was evidently regarded as an aspect of the family

strategy of subsistence, but only as a fall-back tactic, more or less the last resort of a family of good morals to meet an immediate crisis. It is unequivocally clear that the equilibrium between needs and earnings was always taut, as liable to snap through the absence of a bread-earner as through the presence of an additional member in the household. Our Congregation, probably like most outdoor relief organizations, offers us evidence about some cyclical needs. Another crucial aspect of poverty at Florence, which the Congregation failed to confront (presumably because it was not contained in its statutes) and hence which is only hinted at, was the high level of rentals, which explained the overcrowding and cramped habitations of these families. The problem remains to be explored, although the deputies' refusal in one instance to help pay a rent, on the specific grounds that "these cases are too frequent", is an unambiguous indication of its importance.[47] Similarly the whole issue of topographical concentrations of poverty (our contemporary inner-city ghettos) is ignored, although a cursory, impressionistic survey of the applications already indicates numbers of families giving the same address, as well as possibly family networks living in nearby houses.

For the moment these must remain open questions, pointers to other aspects of structural poverty and to the family strategies adopted to confront them. But it is worth recalling that by the eighteenth and nineteenth centuries (and earlier), if outdoor relief was of central importance in western Europe, because of the stress placed on economic independence by society, alternative (and indeed concurrent) strategies of survival were open to the poor, even though they could only have affected significantly smaller numbers. At one extreme, because of the moral taboo on allowing death through visible neglect, families or individuals could opt for a total and socially accepted dependence on institutional support, whether temporary (a period of *renfermement*, the deposit of a legitimate babe at the foundling hospital) or permanent (old-age residence in a hospital). At the other extreme, poor individuals and families could develop an abusive dependence on society by adopting social practices formally and usually juridically unacceptable to that society (such as prostitution, theft or begging). It is evident that the choices were rarely so stark and drastic. Indeed, it is most likely that within the life cycle of the poor, use was made of various alternatives or combinations of them, depending on the nature of the relationships between each individual and the society in which he (or she) recognized himself (or herself); hence the importance of less formal ties of kinship or clientelism. But it is also possible to suggest that strategies of subsistence made more use of institutional resources by the eighteenth-nineteenth centuries precisely because the range and specialization of charitable (and repressive) institutions had widened as a response to a growing awareness of the complexity of society and the incidence of domestic crisis.

## Notes

This study would not have been possible without the earlier assistance provided by the Social Science Research Council, grant HK.6583. I should also like to thank Anna Gozzini.

[1] P. Townsend and D. Wedderburn, *The aged in the welfare state* (London, 1965); P. Townsend, *The concept of poverty* (London, 1970); P. Townsend, *Poverty in the United Kingdom; a survey of household resources and standards of living* (London, 1979); B. Showler and A. Sinfield (eds), *The workless state; studies in unemployment* (Oxford, 1981); A. B. Atkinson, *The economics of inequality* (Oxford, 1983); A. B. Atkinson, A. K. Maynard and C. G. Trinder, *Parents and children: incomes in two generations*, SSRC-DHSS studies in deprivation and disadvantage, 10 (London, 1983).

[2] For a full discussion of the Congregation of San Giovanni Battista and the nature of the sources, see S. J. Woolf, "Charité, pauvereté et structure des ménages à Florence au début du XIXe siècle", *Annales E.S.C.*, 39:2 (1984), pp. 355-382.

[3] Woolf, art. cit., p. 380.

[4] Archivio di Stato di Firenze, Congregazione di San Giovanni Battista, serie Iv, filze 2 to 43. All references and quotations in this article come from this source and will list the *filza* number, civic number of the habitation and surname of the applicant family. Although usually only one or two references are given, they are representative of substantially larger numbers. (Filza) 3: (civic number) 5106 Rossi; 17:5031 Cornamusi; 22:3730 Nannetti, and cf. 19:405 Ricci; 10:3835 Fiorini; 12:5083 Chellini; and cf. 4:4148 Bruni, rejected "because it is a matter of assisting a single individual, of unknown age".

[5] 42:4802 Ferranti; 36:3401 Mochi. One deputy, while recommending the grant of a bed in 1810, noted that a bed had already been granted in 1803: 5:5291 Berretti; while another applicant who had already received two beds in the past 10 years was refused: 2:4917 Aiazzi. The total number of repeated applications (forbidden by the regulations) only amounted to a few dozen.

[6] 6:3941 Socci; cf. 20:4940 Perini ("maintained by the hospital"); 16:4592 Coli ("the serious and dangerous illness of the head of household"). Illness or hospitalization of the head of household was a common justification for granting help, e.g. 18:5342 Agostini; 18:3517 Sonni; 3:3891 Bastianelli; 35:4982 Moschini.

[7] 15:3668 Puliti; 24:3728 Benini ("not of good morals, as not even their children are educated"; 19:4338 Bartolini.

[8] J. M. de Gérando, *Le visiteur du pauvre: mémoire qui a remporté le prix proposé par l'Académie de Lyon sur la question suivante: "Indiquer les moyens de reconnaître la véritable indigence, et de rendre l'aumône utile à ceux qui la donnent come à ceux qui la reçoivent"*, Paris, L. Colas, 1820. An English translation was published in 1833, an Italian one in 1834; by 1844 11 French editions had been published.

[9] 25:3730 Saracini, and many others. There are some rare exceptions, where the visitor's negative recommendation was overturned: 26:3759 Miraceli, 14:3537 Manfriani, 28: 3528 Bernini.

[10] 19:3367 Stiattesi; 22:4020 Nasi.

[11] 2:4709 Papi; 2:5154 Fossi; 4:4136 Borsacchi; 12:4088 Cheli. In one instance (4:4534 Zanoboni) an irritable rapporteur rejected the application on the grounds that the visitor-deputies had recommended a bed, whereas the request was for clothes: "it's impossible to know what is the real need".

[12] 31:4705 Bardini: "the doctor who visited the applicant could find no reason why she cannot feed her daughter"; 32:4623 Cavazzani.

[13] 10:3582 Pampaloni; 25:4292 Giamberini; 10:4860 Chiaranti; 10:3808 Vanini; 2:4709 Bassilichi; 10:388é Corsini; 6:4252 Fioravanti; 3:4835 Lolli; 6:5002 Cavini ("It's not

possible that the applicants do not earn enough"); 6:5021 Buncioni. Many more examples could be given.

[14] Woolf, art. cit., pp. 367-374.

[15] 8:3943 Bianchi.

[16] 16:3389 Corti; 20:3453 Tedeschi; 18:4075 Lanzi; 15:4389 Martini; 18:5342 Agostini, and many others.

[17] 19:4401 Cappelli.

[18] For the total number of rejections, Woolf, art. cit., p. 376. 18:3519 Benvenuti (dowry "refused, as against the regulations"); 12:4130 Nesi (coach-driver training); 19:4157 Bichi; 19:3714 Noccioli (beggars); 22:4215 Santi; 6:5027 Panazzoli (blankets).

[19] In files 31 to 36, of 74 applications for a wetnurse subsidy, only 26 were accepted. The visitor-deputies frequently referred to the mother's inability to feed through lack of nutrition ("incapable of feeding through poor nutrition": 36:4971 Conti). 27:4069 Giorgi ("among their other dire needs, we must also note their total lack of clothing"). 40:3393 Frassinesi (bed).

[20] Particularly hard-nosed rapporteurs seem to be those in files 17 and 27, and a compassionate one in file 21. 37:3512 Boccaccini (wetnurse). 12:4890 Fanti; 6:3769 Pichi (parishes); see also 10:3387 Sbrocchi.

[21] 36:4757 Coloretti, 34:4817 Gabbiani; 35:5038 Santoni (nappies). 3:4042 Ricci; 37:4977 Maglierini (bread). 39:6478 Maori; 18:3442 Guidi; 28:4465 Vestri (numerous children). 28:5075 Spe; 18:3954 Guarnieri; 28:3848 Ferroni; 23:4212 Massai; 18:4916 Giorgetti; 23:4075 Macciani; 10:4786 Albizi (living conditions).

[22] 20:4642 Melani (parish priest); and cf. 20:3609 Manetti. 33:3610 Bellucci (wetnurse subsidy). 15:3735 Fedi (alternative subsidy). 28:5052 Pani (begging). 6:4846 Lucheri (bed sold). 40:3961 Furillazzi; 3:3408 Vallesi; 4:4910 Conti (imminent childbirth). 19:3625 Petrini; 9:3760 Masetti (daughter's bed).

[23] 22:4503 Salvagnini (washerwoman); 27:3773 Galletti ("given that her son is about to abandon her in order to take wife"); 18!3442 Vitali, Bartolini (widow and son-in-law). 18:3892 Lori, 3:4312 Piccini; 21:4902 Lippi (conscription); and cf. 21:5182 Sereni, a window "extremely wretched", because she had sold all her belongings to give some money to her conscripted son who had now been killed in battle.

[24] 15:4389 Martini; 42:3774 Parenti (large families). 6:4847 Galleri (single woman).

[25] 43:3791 Pelatti; 26:3772 Rossi; 18:3568 Rastrelli.

[26] 33:4600 Chini ("sick and incapable, she is allowed to continue living where she is out of charity"); 14:3471 Ghelardini, a family of seven, consisting of mother and two married sons, each with one infant (but all in good health); 26:4375 Mazzetti, a family of six, with the head of household unable to work and the eldest son deranged (*demente*).

[27] 3:3905 Farolfi; 22:5372 Fraschetti (earnings); 25:4409 Donzelli (young couple); 3:4074 Pezzati (couple without children); 6:5021 Buccioni (adolescent); 27:3414 Marchi; 12:4089 Seroni (single young woman); 27:4992 Bellucci (old woman); 27:4397 Ciappi (woman helped by children); 6:5042 Mori; 3:5027 Tafani (beds); 11:4968 Aiazzi (marriage). 12:4328 Falchi (silk weaving); 27:4969 Monelli (washerwoman); 14:3391 (innkeeper); 20:4687 Ganti (servant); and cf. 13:4149 Franchi, considered as a servant even though she was not domestic but "does cleaning jobs" on a casual basis.

[28] 19:3982 Miglia (old woman); 27:4020 Nesi (orphan); 20:3909 Marchiani (abandoned wife); 3:5045 Delli (widowed); 40:3784 Bellebuone (recently widowed, with the additional burden of a married daughter who had fled from her husband); 25:3471 Conti (aged parent); 13:4710 Giannelli; 15:3370 Casati (young children); 20:4951 Pietrai; 11:5007 Bassi (insanity); I have counted 7 cases of *dementi* in the total 4 498. 26:3770 Segalari; 3:4705 Ferroni (illness); 18:5342 Agostini; 16:4592 Coli (illness of head of household); 40:3900 Massai; 3:4801 Ceccherini (large family).

[29] 9:5092 Ciabilli; 5:4444 Piamonti (close to Congregation or parish priest); 5:5000 Pini; 13:3348 Montefiori (recommendations); 26:3977 Bacchini; 7:4796 Galli; 12:4720 Cordenio (army); 6:3800 Marchini; 10:4330 Cantini; 23:4220 Albertini; 15:5137 Guarducci (professional skills); 40:4009 Aretini; 17:4756 Brandi (dismissed clerk); 24:4641 Vannetti; 19:4101 Baccini (good birth); 43:4806 Renard (noble).

[30] 3:3857 Borla ("as these two children are of vivacious temperament"); 42:3759 Sacchi ("with a son who lives in the most total idleness, as he has no one to guide him"); all accepted into the Casa Pia.

[31] 10:3473 Landucci; 14:3655 Morosi.

[32] 21:3866 Bartolini ("but someone should keep check that the bed is not sold"); 14:3835 Scarlini; 11:5183 Covoni; 39:3925 Bianchi; 28:5066 Carotti ("as the parents are young"); 6:5042 Mori.

[33] 11:3413 Laghi; 6:4866 Paladini; 6:4897 Mori, 22:4859 Lascialfare.

[34] 40:3748 Mazzoni; 7:4085 Accioli; 4:4245 Belli ("as the sum is substantial", the deputies decided that only items "of essential needs" should be redeemed).

[35] Woolf, art. cit., pp. 367-374.

[36] 33:3773 Bronchelli; 43:5016 Bettini; 20:4240 Cioni.

[37] Woolf, art. cit., pp. 360-362, 364-365.

[38] 16:3389 Corti; 23:3654 Giorgi; 12:3634 Picchi; 12:3833 Bianchi; 4:4449 Gucciardi; 14:3471 Ghelardini; 3:4611 Stefanelli (son's family of four); 43:3957 Martinelli (mother-in-law in another house).

[39] 28:5007 Galli; 6:4866 Paladini; 6:4897 Mori; 6:5033 Frangini.

[40] 12:4909 Gamberucci; 11:3413 Laghi.

[41] 23:4956 Pichianti (where the family insisted on a separate bed as the condition for keeping the orphan; the Congregation agreed, so long as it was her property); 21:4737 Giorgi, 27:4704 Giolli.

[42] 5:5094 Mazzaranghi.

[43] 14:3542 Pestellini; 33:4600 Chini.

[44] For example, see the obscure explanation in 6:4861 Ferroni, where a bed was requested to separate the sister from a "cohabitant".

[45] For the redemption of pawned possession, see above, note 34. 22:4062 Teresa Grifoni ("very small weekly earnings [from spinning] which oblige her to fast on occasion").

[46] Woolf, art. cit., pp. 379-380.

[47] 27:4101 Meini.

# Vom republikanischen Revolutionär zum europäischen Patrioten

WERNER MAIHOFER
Bielefeld/Florenz

## Einleitung: Karl Hillebrand, geboren am 17. September 1829 in Gießen, gestorben am Oktober 1884 in Florenz

Es ist die Rede von Karl Arnold Hillebrand, einem zu Unrecht Vergessenen. Ein Leben und Werk heute mehr denn je des Erinnerns und Bedenkens Wert. Vergegenwärtigen wir uns einleitend einige Stationen dieses Lebens, bevor wir mit der Evolution des Werkes uns beschäftigen, das aus ihm hervorgeht[1].

Karl Hillebrand wächst auf in einem freisinnigen Elternhaus. Der Vater ist Professor der Philosophie in Gießen, weitbekannt durch seine dreibändige "Deutsche Nationalliteratur seit dem Anfang des 18. Jahrhunderts". Der Sohn beginnt nach der Schulzeit, begeistert von den Schriften Proudhons (La propriété, c'est le vol), ein Studium der Jurisprudenz, als 1848 die Revolution ausbricht, an der er bis zum bitteren Ende der endgültigen Niederlage auch der badischen Revolutionstruppen 1849 teilhat, mit denen er in der Festung Rastatt gefangengesetzt wird, wo er täglich seine Erschießung erwartet. Mit Hilfe der Schwester gelingt ihm die Flucht aus der Festung im Kugelregen der Verfolger über den Rhein nach Frankreich.

In Paris wird Hillebrand Sekretär Heinrich Heines, der ihm seinen Romanzero in die Feder diktiert. Hillebrand wird tief berührt vom heiligen Ernst schöpferischer Arbeit, dem er hier erstmals begegnet. Nicht zuletzt auf Heines Rat, für den er ihm ein Leben lang dankbar bleibt, entschließt Hillebrand sich, seine akademische Karriere in Frankreich fortzusetzen. Er geht schon 1850 nach Bordeaux, wo er nach harten Anfängen als Deutschlehrer und Studentenaufseher 1854 ein Studium der Philologie, der Historie und der Literatur aufnimmt, das er 1858 mit der Licence-ès-Lettres ehrenvoll abschließt.

Auch als in diesem Jahr die Verurteilten von 1849 begnadigt werden, bleibt Hillebrand in Frankreich und setzt sein Studium in Paris fort, wo er

1861 an der Sorbonne mit einer Dissertation über den Florentiner Historiker Dino Compagni promoviert, mit der er sich in der Fachwelt einen Namen macht. Nach Bordeaux zurückgekehrt legt er 1863 auf die Ausschreibung der dortigen Akademie eine Schrift: "De la condition de la bonne comédie" vor, die preisgekrönt wird. Noch im selben Jahre wird er auf die Professur für ausländische und vergleichende Literatur an die Universität von Douai berufen.

Er gibt aus seinen eigenen Vorlesungen einen ersten Sammelband "Etudes Italiennes" heraus, der Essays über Dante, über Machiavelli, über Savonarola enthält, die auch in Italien begeisterte Aufnahme finden. Es öffnen sich im die berühmtesten Zeitschriften seiner Zeit, durch die er mit historischen und literarischen Essays einer größeren Öffentlichkeit in Frankreich, in Italien, aber auch in Deutschland und England bekannt wird. Nicht zufällig nimmt Hillebrand 1865 als offizieller Vertreter Frankreichs an der Gedenkfeier in Florenz teil. In den folgenden Jahren bereist er im Auftrag des französischen Unterrichtsministers Deutschland, Holland, Belgien und die Schweiz, um über das Schulwesen dieser Länder zu berichten. Seinen umfassenden Bericht "De la Réforme de l'Enseignement Supérieur" folgt 1868 ein ergänzender Aufsatz über "L'instruction publique en Italie". Da zerschlägt der Deutsch-Französische Krieg, mit dem Hillebrand "einen lieben, geistreichen Freund plötzlich dem finstersten Wahnsinn verfallen" sieht, jede weitere Möglichkeit geistigen Wirkens im Verhältnis zwischen Frankreich und Deutschland. Er beschließt, in England Zuflucht zu suchen. Fast wird er auf der Fahrt in Lille als vermeintlicher preußischer Spion gelyncht.

In England lädt die Times Hillebrand ein, im deutschen Lager über den Fortgang des Krieges zu berichten. Er lehnt diese Einladung ebenso ab, wie das Angebot der preußischen Regierung, im Elsaß das deutsche Schulwesen einzuführen. Stattdessen geht er als Korrespondent der Times in die provisorische Hauptstadt Italiens: Florenz, um im Zug der Truppen bis nach Rom über den Fortgang der Einigung Italiens zu berichten, an der er mit ganzem Herzen sich beteiligt.

Von 1871 an kehrt Hillebrand mehrfach nach England zurück, um Material für seine "Geschichte Frankreichs", die ihn bis an sein Lebensende beschäftigt, zu sammeln. Auch als er sich 1873 endgültig in Florenz niederläßt, um als freier Schrftsteller zu arbeiten, setzt er seine Besuche in Deutschland und England fort. So ist er 1879 persönlicher Gast Gladstones und der englischen Whigpartei und hält auf Einladung der Royal Institution of Great Britain englische Vorträge über deutsche Geistesgeschichte, deren Veröffentlichung ihm auch in England einen Namen macht.

Karl Hillebrand heiratet im selben Jahre Jessie Taylor "nach 28 Jahren treuer Freundschaft", der er schon im Sommer 1850 in Bordeaux als Gattin eines Weinhändlers begegnet, mit dem sie seit ihrem 16. Lebens-

jahr in unglücklicher Ehe lebt, wenige Monate, nachdem Richard Wagners Versuch scheitert, mit ihr in den Orient zu fliehen. Jessie trennt sich bereits 1853 von ihrem Mann und lebt in Dresden und Stuttgart, später in Rom und Florenz. Eine hochgebildete Frau, wohlvertraut mit englischer, französischer, wie mit deutscher Literatur, deren mit dem Vater Hillenbrands befreundeter Vater eine bahnbrechende übersetzung deutscher Minnesänger verfaßt. Aber sie ist vor allem auch, was Karl Hillebrand gänzlich abgeht, eine begabte Musikerin, die mit Hans von Bülow, Peter Cornelius und anderen sich befreundet, für Liszt, Joachim und viele andere in Italien Konzerte organisiert und selbst die Donnerstagabendkonzerte der von ihr 1860 in Florenz gegründeten Società Cherubini dirigiert.

Mit Jessie wird in Karl Hillebrands letztem Lebensjahrzehnt, das reiche schriftstellerische Ernte bringt, das gemeinsame Haus am Lungarno, das er schon von 1875 an mitbewohnt, zum Mittelpunkt eines gesellschaftlichen Lebens, das "das ganze gebildete Europa umfaßt", wie Hillebrand selbst sagt. In dem so die literarischen, musikalischen und künstlerischen Freunde aus all den Stationen beider Lebenswege durch die "Zeiten, Völker und Menschen" sich einfinden, von fern und nah. Und in dem zuletzt selbst Richard und Cosima Wagner auf ihren Italienbesuchen einkehren.

Man hat von Karl Hillebrand gesagt, er sei ein "Genie der Sympathie" gewesen. Dies gilt nicht weniger für Jessie Hillebrand. Beide sind das gewesen, was Hillebrand einmal von Montesquieu schon in Bordeaux sagt: "In die Freundschaft verliebt" (amoureux de l'amitié).

Karl und Jessie Hillebrand sind beide in Florenz verstorben. Sie liegen, umgeben von Freunden, vereint auf dem Friedhof Agli Allori vor den Toren von Florenz begraben.

Schon als *republikanischer Revolutionär* aus großbürgerlichem Gelehrtenhause ist Karl Hillebrand, der in Frankfurt, wie Richard Wagner in Dresden, auf die politische Barrikade geht, von Anfang eine *beispielhafte Gestalt* dieses gescheiterten Aufstandes, dem in unserer Zeit, dem in unserem Lande einer endlich erreichten Republik, als einem frühen Republikaner geschichtliche Aufmerksamkeit und ehrendes Gedächtnis sicher sein müßten. Doch daß aus diesem republikanischen Revolutionär Karl Hillebrand am Ende, nach erfolgter Einkerkerung, drohender Hinrichtung, abenteuerlicher Flucht, im Exil in Frankreich, danach in England und zuletzt in Italien etwas geworden ist, was es damals so eigentlich nicht geben konnte und noch heute kaum gibt: ein *europäischer Patriot*, verleiht der Figur Hillebrands für unsere Zeit und Welt etwas Exemplarisches, über die seine und noch die unsere hinaus.

Denn hier entsteht auf einem außergewöhnlichen Lebensweg und geistigen Entwicklungsgang ein Mensch, der nach und nach das ganze Europa als sein erstes, zweites, drittes und viertes Vaterland begreift und empfindet, und der so, in der Hochzeit des Nationalismus mit seinem

unbeirrbaren Bemühen um vorbehaltslose Unparteilichkeit als eine ganz und gar *unzeitgemäße Gestalt* erscheint; der den "Franzosen zu deutsch" und den "Deutschen zu französisch" und am Ende beiden zu englisch oder gar zu italienisch: zu europäisch denkt und fühlt.

Was schon Nietzsche mit scharfer Witterung für das Unzeitgemäße in seiner Zeit in einem Briefe an Hillbrand vom Frühjahr 1878 zu der hellsichtigen Bemerkung veranlaßt: "Nach einem Winter schwerer Erkrankung genieße ich jetzt im Wiedererwachen der Gesundheit Ihre vier Bände 'Völker, Zeiten und Menschen' und freue mich darüber wie als ob es Milch und Honig wäre. O Bücher, aus denen eine europäische Luft weht, und nicht der liebe nationale Stickstoff! Wie das den Lungen wohltut!" Und dann: "Ich möchte den Autor sehen, der Ihnen in Unbefangenheit und wohlwollender Gerechtigkeit gleichkäme." Von daher ist Karl Hillebrand die geschichtliche Aufmerksamkeit als einer der ersten wahrhaften Europäer in der Vorgeschichte der europäischen Idee gewiß. *Vom republikanischen Revolutionär zum europäischen Patrioten* so lautet darum unsere Umschreibung des Eigentümlichen und Unverwechselbaren, das diesen Karl Hillebrand ausmacht und auszeichnet, wenn wir diese Person begreifen und verstehen in ihren politischen Ideen als Phänomen einer historischen Epoche; als ein Exemplum dessen, was in dieser Zeit und Welt anhebt, zugleich als ein Novum dessen, was über eben diese Zeit und Welt hinausweist. Lebendige Wiederlegung des apodiktischen Diktums Hegels, daß wie schon die Philosophie nur "ihre Zeit in Gedanken erfaßt", auch das Individuum "als Sohn seiner Zeit" niemals über den Geist seiner Zeit hinauszudenken vermöge.

Dennoch ist diese geistige Entwicklung Karl Hillebrands, vom gescheiterten republikanischen Revolutionär zum verfrühten europäischen Patrioten, die sich aus dem äußeren Werdegang und schriftlichen Lebenszeugnis ablesen lässt, alles andere als aus sich heraus einsichtig und von selbst verständlich.

Die Frage ist nicht nur, ob und wie die eine mit der anderen Seite dieser Gestalt zusammenhängt, oder gar aus ihr hervorgeht? Die Frage ist schon, was wird aus dem republikanischen Revolutionär, der nur durch die glückliche Flucht der sicheren Erschießung als Märtyrer einer Revolution entgeht? Ein *enttäuschter Revolutionär*, der sich in eine apolitische Existenz ästhetischer Reflexion zurückzieht? Oder ein *unbeirrter Republikaner*, der auf seinem inneren Weg durch die "Zeiten und Völker" eben der politischen Idee treu bleibt, die ihm von Anfang bis Ende über alles geht: der "Freiheit des Individuums"? Ein *verschreckter Konservativer* also, der sich in seiner politischen Existenz und literarischen Produktion mit der zeitgenössischen Restauration verbindet? Oder ein *entschiedener Reformer*, der die Verwirklichung der Freiheit des Einzelnen über alle zeitgenössischen politischen Forderungen nach Gleichheit, ja selbst nach Brüderlichkeit stellt? Und dem darum der Vorrang der Freiheit auch über alle

zeitgenössischen nationalen Forderungen nach Einheit geht? Und der so im Zeitalter der Restauration und des Nationalismus als republikanischer Revolutionär nicht zufällig, sondern schicksalhaft zum europäischen Patrioten wird?

Wir wollen versuchen, uns den Antworten auf diese Fragen nicht mit pauschalen Mutmaßungen über die politischen Ideen Karl Hillebrands anhand der politischen Stationen seines *Lebens* zu nähern, über deren entscheidende erste: Gießen wir ebensowenig Verläßliches wissen, wie über deren nicht minder entscheidende zweite: Bordeaux. Wir wollen vielmehr versuchen, sein *Werk* selbst, das uns in sieben Bänden unter dem Titel: "Zeiten, Völker und Menschen"[2] vorliegt, daraufhin zu befragen, was sich in ihm an *exemplarischer Evolution der politischen Ideen* vollzogen hat, vom republikanischen Revolutionär der 48er Bewegung am Anfang bis zum europäischen Patrioten am Ende seines reichen Lebenswerkes.

## 1. Der liberale Republikaner

Wie die Debatten der Paulskirche um das Thema der Freiheit zeigen, geht es dieser Bewegung bis auf vereinzelte Stimmen weder um eine sozialistische Revolution der Gleichheit im Sinne der zweiten Phase der französischen Revolution von 1793, mit der auch in der Verfassung die Gleichheit vor die Freiheit rückt, noch auch nur um eine demokratische Revolution im vollen Sinne der ersten Phase der französischen Revolution von 1789, die eine Gesellschaft der "Freiheit, Gleichheit, Brüderlichkeit!" auf ihre trikoloren Fahnen schreibt[3].

Es geht allein um die Freiheit des Einzelnen und seine Stellung im Staat, die als eine solche gleicher Freiheit und zugleich gesetzlicher Gleichheit eingefordert wird. Wo überhaupt von Gleichheit, ist so hierbei allenfalls von Gleichheit vor dem Gesetz, von Gleichbehandlung durch den Gesetzgeber die Rede.

Auch Karl Hillebrand spricht als ein *liberaler Republikaner* in diesem Sinne sein Leben lang von einem Staate, der sein Ziel nicht in sich selbst, sondern der die "Freiheit des Individuums" zum Zwecke hat (III, 121). Ein solcher auf *Liberalität*: den Vorrang der Freiheit des Einzelnen begründeter und gerichteter Staat heißt schon für Rousseau wie später für Kant *Republik*, wenn diese Ordnung der Freiheit aller durch eine Herrschaft der Gesetze aus dem allgemeinen Willen zum gemeinsamen Nutzen gesichert ist. Für den Republikaner kommt es für die Rechtmäßigkeit einer Herrschaft darum nicht auf die sog. Staatsform an: die Monarchie, Aristokratie oder Demokratie, sondern darauf, ob in ihr zwischen Regierenden und Regierten "Interessen- und Willenseinheit" besteht, wie schon Rousseau sagt. Ob also die Herrschaft des Gesetzes Ausdruck und Werkzeug des besonderen Interesses eines einzelnen oder einiger Weniger, oder des

gemeinsamen Interesses aller ist. Insoweit ein Staat dies erreicht: res publica, "gemeinsame Sache" aller Bürger zu sein, verdient er unabhängig von seiner Form den Namen Republik. Insofern kann schon für Rousseau wie noch für Hillebrand folgerichtig auch "eine Monarchie Republik sein", wo sie von einer Herrschaft des Gesetzes aus dem allgemeinen Willen zum gemeinsamen Nutzen bestimmt ist; aber auch eine bloße Despotie, wo in ihr der partikulare Wille und das besondere Interesse des Monarchen vorherrscht.

In eben diesem Sinne heißt es bei Karl Hillebrand in einem Vergleich der Monarchie Friedrichs des Großen mit der Napoleon Bonapartes höchst bezeichnend: "Friedrich nannte sich vom Tage seiner Thronbesteigung an den ersten Domestiken des Staates und er handelte bis zum letzten Atemzuge nach diesem Grundsatz. Das erste Wort des Jünglings an die Staatsbeamten ging dahin, daß sie keinen Unterschied zwischen König und Staat machen dürften und, wenn beide Interessen je kollidieren sollten, sie das Staatsinteresse vor dem Interesse des Königs zu wahren hätten. . . Und auf seinem Sterbebette, nach sechsundzwanzig Jahren einer glorreichen Regierung, empfahl er nicht als oberste Regel seinem Nachfolger und allen seinen Verwandten 'immer ihren persönlichen Vorteil dem Wohle des Landes und dem Vorteil des Staates zu opfern'? Wir wissen, daß das keine Worte waren. Für Bonaparte dagegen war der Staat nie etwas anderes als er selber. . . Hatte Kant gelehrt, man solle jeden Mitmenschen stets als Zweck ansehen, so predigte Napoleon durch die Tat, daß man sie nur als Mittel brauchen dürfe. Nie ist die Menschenbenutzung im eigenen Interesse zu einer größeren Virtuosität gebracht worden" (III, 283ff.).

Karl Hillebrand bleibt auch in solchen zunächst befremdlich erscheinenden Äußerungen somit *unbeirrter Republikaner* aus derselben Grundüberzeugung, die schon Rousseau und Kant als Republikaner kennzeichnen, für die nicht die Staatsform: Monarchie, Aristokratie oder Demokratie über die Rechtmäßigkeit einer Herrschaft von Menschen über Menschen entscheidet, sondern das Herrschaftsprinzip: Herrschaft des Gesetzes aus dem allgemeinen Willen zum gemeinsamen Nutzen zu sein, oder: in der einen wie der anderen Form, gemessen an diesem Prinzip, nicht zu sein.

Mit dieser Option für die politische Idee der Republik verbindet sich bei Karl Hillebrand schon früh die Option für die politische Idee der Reform. Nicht einfach als ein durch das Erleben und Erleiden des gescheiterten Aufstandes 1848 "enttäuschter Revolutionär", sondern als ein durch die nachherige Erfahrung der "französischen Zustände", aber auch das lebenslange historische Studium der französischen Revolution durch die Geschichte belehrter, *entschiedener Reformer.*

In Absetzung von der extremen Position Burkes stellt auch Karl Hillebrand fest: "Wohl war die große Revolution notwendig, das muß

zugegeben werden, und keine Menschenkunst hätte sie aufhalten können". Aber auch er fragt sich: "Mußte wirklich all' dies Blut vergossen werden, um die neuen Zustände zu schaffen? Das Beispiel von P. Leopolds Wirksamkeit in Toscana dürfte das Gegenteil beweisen" (V, 62).

In Übereinstimmung mit Taine sieht auch Karl Hillebrand "die wahre innerste Natur der großen Bewegung", aus der letztlich der revolutionäre Impuls hervorgeht im Umsturz der Eigentumsverhältnisse. "Was auch die großen Namen Freiheit, Gleichheit, Brüderlichkeit sein mögen (so sagt er mit Taine), mit denen die Revolution sich schmückt, sie ist in ihrem Wesen *eine Verrückung des Eigentums*. Darin besteht ihre erste Triebfeder, ihre dauernde Macht, ihr wahrer Halt, ihre historische Bedeutung. Einst, im Altertum hat man dergleichen gesehen, Schuldentilgung, Güterkonfiskation, Teilung der Gemeinde- und Staatsgüter, aber die Sache beschränkte sich auf eine Stadt, auf ein kleines Gebiet. Zum ersten Male wurde sie nun in einem modernen Großstaat vollzogen" (V, 199). Auch in ihm bleibt es für Hillebrand dabei: "Auch dieses Verhältnis hätte man ruhig und friedlich durch Ablösung regeln können, wie in Preußen im Jahre 1808, in Rußland im Jahre 1861" (V, 194).

Karl Hillebrand bleibt bis zu seinem Lebensende im italienischen Risorgimento dieser entschiedene Reformer, der selbst eine solche grundstürzende Veränderung der Verhältnisse, wie die Eigentumsverhältnisse, auf friedlichem Wege für möglich hält. Dennoch läßt sich diese seine zunächst in ihren positiven Konturen beleuchtete politische Position eines unbeirrbaren Republikaners und entschiedenen Reformers, auch in ihren negativen Konsequenzen nur erhellen und verstehen aus zwei weiteren, damit verbundenen politischen Optionen, die zum Abschluß unserer Würdigung der beispielhaften Gestalt Karl Hillebrands als eines Republikaners im Zeitalter der Reaktion zur Sprache kommen sollen: die für den Konservatismus und gegen die Reaktion, und die für den klassischen gegen den modernen Liberalismus. Sie lassen uns Hillebrands Gestalt beim zweiten Zusehen als eine höchst differenzierte und komplizierte politische Figur erscheinen: nicht eines reaktionären, sondern eines *reformerischen Konservativen*, aber auch: nicht eines demokratischen, sondern eines *republikanischen Liberalen*. Die erste dieser beiden politischen Optionen wird deutlich in Hillebrands Auseinandersetzung mit der Gestalt Metternichs: dem reaktionären Konservativen, die zweite in seiner Absetzung vom Denken John Stuart Mills: des demokratischen Liberalen.

Hillebrand sieht Metternich als eine für seine Zeit und Welt verhängnisvolle Gestalt: als einen Revolutionär unter umgekehrten Vorzeichen, somit nicht als einen Konservativen, sondern als einen Reaktionär. Aus der Auseinandersetung mit dieser politischen Figur wird, wie nirgendwo sonst, schlagartig deutlich, was für Hillebrand die eigene politische Option für den Konservatismus und gegen die Reaktion meint und heißt. "Die Reaktion blieb sein politisches Ideal (sagt er von Metternich); und er

glaubte konservativ zu sein, wo er nur ein *umgekehrter Revolutionär* war. Der Grundirrtum der festländischen Politiker beider entgegengesetzten Lager, die noch immer Reaktion und Konservatismus identifizieren und überdies die Kirche als notwendigen Verbündeten der konservativen Interessen ansehen, ward so recht von Metternich und seiner Schule eingeführt. Der wahre *Konservative* hat einen zu festen Glauben in die erhaltenden Kräfte der Gesellschaft, um ihnen durch gewaltsame Reaktion zur Hilfe zu kommen. Ihm scheint Aberglauben und Priesterherrschaft eine größere Gefahr für den Staat und seine ruhige Entwicklung, als Freiheit und Öffentlichkeit, welche ja die einzige Atmosphäre für gesundes, normales Leben sind. Für den *Reaktionär* ist künstlicher Stillstand, womöglich künstliches Zurückzwängen der Zustände, ist künstlich erhaltene Heimlichkeit und Dunkel und Schweigen die Summe aller Staatskunst und die Lebensluft ihrer Tätigkeit. *Unbeschränkte Freiheit erschreckt den Konservativen nicht, wenn nur die Herrschaft des Gesetzes nie in Frage kommt*; das Reden und Schreiben der Laien läßt er gewähren, solange nur das Handeln den Sachverständigen allein gewahrt bleibt; *der Umwandlung der Verhältnisse setzt er keinen Damm entgegen, nur dem Umsturze*; wie er auch nicht die Änderung der Gesetze nach Zeit und Umständen, sondern nur die Gesetzgebung nach aprioristischen Theorien bekämpft. *Der Reaktionär im Gegenteil gleicht dem Revolutionär in seiner Vorliebe für solche Theorien, für gewaltsame Herstellung gewisser Zustände, in seiner Unduldsamkeit für die Meinung anderer*" (V, S. 342f; Hervorhebungen von mir).

Dieses konservative Manifest im Zeitalter der Reaktion spricht für sich selbst und es spricht für Hillebrand als das, was er gegen den Strich seiner Zeit beispielhaft ist: ein *reformerischer Konservativer*, der im Gegensatz zu Burkes "modernem Vorurteil", "daß politischer und religiöser Konservatismus zusammen gehen müssen", einen wie er sagt: "höheren Konservatismus" vertritt, der *wenigstens so viel Skepsis voraussetzt, als zur Toleranz nötig ist* (V, S. 67); einen wie man auch sagen könnte: aufgeklärten Konservatismus, der über seine Zeit einer noch oder wieder geheiligten Allianz von "Thron und Altar" hinausweist.

Sehr viel undeutlicher und zwiespältiger, aber auch zeitgemäßer, wenn nicht zeitbedingter, erscheint uns die Gestalt Karl Hillebrands, wenn wir über sein stetes Bekenntnis zur Freiheit als dem höchsten Zweck des Staates hinaus, nach seinem Verhältnis fragen zu den Forderungen der Gleichheit oder der Brüderlichkeit, die doch schon in der ersten Phase der demokratischen Revolution in Frankreich nebeneinanderstehen und zusammenklingen; oder gar nach seinem Verständnis von Demokratie, von Liberalismus und Sozialismus als politischen Ideen, von denen doch seine Zeit und Welt voll ist.

Eine erste Antwort auf diese Fragen ergibt sich aus der Auseinandersetzung: der teilweisen Zustimmung, aber auch der entschiedenen Absetzung Karl Hillebrands von John Stuart Mill, dem bedeutendsten Denker

des Liberalismus seiner Epoche, dem Vorläufer des über den klassischen Liberalismus hinausführenden modernen: demokratischen und sozialen Liberalismus.

Solange und soweit Mill "über die Natur der Freiheit spricht", hält Hillebrand diese Theorie für "durchaus unantastbar" (III, s. 120), und darum Mills Versuch "über die Freiheit" für das Beste was er geschrieben hat (III,115). Ausdrücklich macht sich Hillebrand nicht nur diese Theorie als politische Idee zu eigen mit den Worten: "Den Menschen, hatte Mill gesagt, solle nur zum Selbstschutz erlaubt sein dürfen, die Freiheit ihrer Mitmenschen zu beschränken; Gewalt würde nur dann mit Recht über ein Mitglied der Gesellschaft gegen dessen Willen ausgeübt, wenn es sich darum handle, zu verhindern, daß andern ein Leid angetan werde, oder mit drastischeren Worten: 'Laß jeden tun, was ihm beliebt, so lange er seinen Nächsten nicht verletzt'" (III, S. 122). Ebenso ausdrücklich hält er diese Theorie auch der Anwendung auf die Praxis für fähig, wenn er sagt: Die Freiheit "Wie Mill sie definiert, ist bis zu einem gewissen Grade der Verwirklichung fähig und wird es von Tag zu Tag mehr werden" (III, 120). Aus dieser "absoluten Freiheit" des Einzelnen aber folgt auch für Hillebrand wie für Mill, daß "die Selbstverteidigung allein" den "Staat berechtigt, gegen die Freiheit des Individuums einzuschreiten", um so die Selbstsucht der Menschen "in den Grenzen zu halten, wo sie die Rechte und Freiheit anderer nicht beeinträchtigt" (III, 120f.).

So entschiedene Zustimmung zu diesem von Mill definierten Ideal des *klassischen Liberalismus* Hillebrand äußert, so entschieden ist seine Ablehnung der damit bei Mill erstmals verbundenen anderen politischen Postulate der demokratischen Revolution nach "Freiheit, Gleichheit, Brüderlichkeit", schon in ihrer ersten Phase 1789. Hillebrand erklärt dazu rundweg: "Gleichheit und Brüderlichkeit sind zwei Ideen, wenn man will: zwei Ideale, welche dem Menschen absolut widernatürlich sind; denn das ganze Dasein der Menschheit beruht ja gerade auf Ungleichheit und Selbstsucht; man könnte sich dieselbe gar nicht denken, ohne den 'Kampf ums Dasein', und es bedurfte kaum des Darwin'schen Gesetzes, um denkende Beobachter, Historiker oder Politiker davon zu überzeugen" (III, S. 120). Carl Hillebrand verharrt bis zum Ende seines Lebens, wie der in Band VII von seiner Frau Jessie herausgegebene Nachlaß überdeutlich zeigt, in dieser politischen Position des *liberalen Republikaners*, dem schon Mills Option für die *Demokratie* als "die wesentlichste Garantie für gute Regierung" (III, 112) nicht mehr nachvollziehbar ist; und der gar "Mills Teilnahme an der Agitation gegen die bestehenden Agrarverhältnisse" verdächtigt, "im Dienste des *Sozialismus* zu stehen" (III, 115).

Zutiefst ist gerade in diesen späten Äußerungen die Erschütterung spürbar, die selbst einen so "freien Geist" wie Karl Hillebrand angesichts der grundstürzenden Veränderungen ergreift, mit denen im heraufkommenden Zeitalter der Industriegesellschaft und späteren Massendemokra-

tie die auf Vorrechte des Besitzes oder doch der Bildung gegründeten *Ständegesellschaften* unaufhaltsam sich auflösen auch da, wo mit der monarchischen Restauration gegen die demokratische Revolution, der Feudalismus über dem schon einsetzenden Kapitalismus noch eine Zeitlang fortlebt.

Hillebrand erkennt sehr wohl, daß die "neue Weltanschauung", die aus dem *französischen Rationalismus* hervorgeht, als "in Frankreich die Literaten gegen die Mitte des 18. Jahrhunderts die Hauptpolitiker wurden" (wie er mit Tocqueville sagt), eine "tiefgehende Umwälzung zur Folge gehabt" hat, ohne jedoch das "Wesen des Staates und der Gesellschaft zu ändern". "Wohl sind die Ansprüche des Rationalismus, die Gesetzgebung nach vorgefaßten Begriffen von Gleichheit und Gerechtigkeit zu regeln, nicht ohne Einfluß geblieben. . . Wohl ist in den meisten Ländern jeder Staatsbürger als gleichberechtigt und gleichwertig anerkannt worden; tatsächlich aber ist die Gewalt in den Händen der Gebildeten und Besitzenden geblieben" (VII, S. 186 ff).

Hillebrand anerkennt auch sehr wohl, daß diese der *modernen Philanthropie* zugrunde liegende Weltanschauung "auch manches ins Leben gerufen hat, das die Existenz der arbeitenden Klassen innerhalb ihres Standes erleichtert und gebessert, sie in Krankheit, Alter und Arbeitslosigkeit unterstützt, ohne ihre normale Existenz durch trügerisches Vormachen besserer Zustände zu verderben". Aber er meint ernstlich: "die ganze große Menschheitsmaschine würde stille stehen, wenn wir uns fortwährend an die Stelle anderer setzen und uns abmühen wollten, den Forderungen einer abstrakten Gleichheit entsprechend, allen dieselben Lebensbedingungen zu sichern. So bleibt's denn auch bei frommen Wünschen, *genügend, die Menschen*, die früher ohne viel nachzudenken ganz glücklich in ihrer beschränkten Existenz waren, *mit ihrem Los unzufrieden zu machen; nicht genügend dieses Los zu ändern*" (VII, S. 186 f.; Hervorhebungen von mir).

Hillebrand fühlt so sehr wohl den inneren Widerspruch seiner Zeit und Welt: des in ihr aufkommenden Bewußtseins der Gleichheit und der in ihr fortbestehenden Verhältnisse der Ungleichheit; also von "frommen Wünschen" nach Gleichheit, die zwar "genügend" sind, die Menschen mit "ihrem Los unzufrieden zu machen", aber "nicht genügend" sind, "dieses Los zu ändern". Es ist eben diese Halbheit, so oder so, die diesen Widerspruch, die solche Unzufriedenheit hervorruft und verstärkt. Was uns hier an der beispielhaften Gestalt Karl Hillebrands als eines liberalen Republikaners von Anfang bis Ende begegnet, ist mehr als nur das persönliche Temperament einer Person, es ist die geschichtliche Atmosphäre dieser Epoche: der "Gründerzeit" des Industriekapitalismus, der selbst das Erzeugnis und Zeugnis einer allenfalls zur Hälfte verwirklichten Republik war und ist.

Es ist nicht nur bezeichnend für den Geist einer Zeit, welche großen

Gedanken in ihr geschichtsmächtig werden; es ist ebenso bezeichnend, für welche großen Gedanken, die am geistigen Ursprung einer Epoche "zum Vorschein" kommen, "zur Sprache" gelangen, diese Zeit bis zum Ende geschichtsblind bleibt. Auch hier kann die Person Hillebrands als ein Exemplum stehen dafür, was von der ursprünglichen *politischen Idee der liberalen Republik bei Rousseau und Kant* in der Folgezeit fortwirkt, aber auch fortfällt.

Hillebrand geht, wie Rousseau und Kant davon aus, daß der Mensch tatsächlich von "Ungleichheit und Selbstsucht" bestimmt ist, wie wir sahen; also von dem, was Kant die "ungesellige Geselligkeit" genannt hat, deren Widerstreit und Wettstreit nur in dem gesetzlichen Gehege einer bürgerlichen Gesellschaft so in Grenzen gehalten werden kann, daß sich der Mensch nicht in "wilder Freiheit" selbst zerstört. Ein "Antagonismus" der "Ungeselligkeit": eine *Dialektik der Freiheit,* aus der für Kant alle menschliche Entwicklung der Naturanlagen und Geisteskräfte hervorgetrieben wird, und die so nur in einer gesetzmäßigen Ordnung größtmöglicher und gleichberechtigter Freiheit und Sicherheit zur bestmöglichen Geltung und Wirkung gebracht werden kann, die wir heute einen *freiheitlichen Rechtsstaat* nennen[4].

Zu alledem finden sich Anklänge auch im Denken Hillebrands. Jedoch ist in den vielfältigen Hinweisen auf Kant jede Erinnerung daran verloren, daß schon für diesen Denker der liberalen Republik *Freiheit und Gleichheit* zur Begründung der *Stellung des Bürgers im Staate* unauflöslich zusammengehören, als das "angeborene Recht der Freiheit und Gleichheit" kraft der "Menschheit in jedermanns Person", wie er ausdrücklich sagt. Ebenso ist in den zahlreichen Bezugnahmen auf Rousseau die übereinstimmende Einsicht nicht wiederzufinden, daß nicht nur ohne die *Gleichheit der Freiheit des Bürgers im Staate* auch "die Freiheit keinen langen Bestand haben kann", sondern daß eine liberale Republik zugleich auch eine angemessene und verhältnismäßige *Gleichheit der Wohlfahrt der Bürger in der Gesellschaft* gewährleisten muß. Deshalb erklärt es schon Rousseau für eine der "wichtigsten Aufgaben der Regierung", "einer übermäßigen Ungleichverteilung der Güter vorzubeugen, nicht indem sie den Reichen ihren Besitz entzieht, sondern allen die Mittel nimmt, Reichtum anzuhäufen; nicht indem sie Armenhäuser baut, sondern die Bürger vor Verarmung bewahrt".

Was Rousseau hier schon im ersten Ansatz seiner politischen Idee einer liberalen Republik für die "wahre Wirksamkeit eines Staates erklärt": "allmählich alle Vermögen jener Ausgewogenheit anzunähern", ist nichts anderes als die Umschreibung dessen, was wir heute einen *freiheitlichen Sozialstaat* nennen[5].

Es ist höchst bezeichnend für den Geist jener Zeit und Welt, daß selbst für einen so unabhängigen Kopf wie Karl Hillebrand: den *reformerischen Konservativen und republikanischen Liberalen,* beide mit der origina-

len Idee einer liberalen Republik verbundenen Ziele des Staates, die wir heute als die eines *freiheitlichen Rechtsstaates und Sozialstaates* bezeichnen, dem geschichtlichen Vergessen verfallen sind. Auch darin wird uns diese Person Karl Hillebrand, in der exemplarischen Evolution ihrer politischen Ideen, als das Phänomen einer bestimmten historischen Epoche sichtbar, als eine beispielhafte Gestalt und zugleich unzeitgemäße Gestalt, auf dem geistigen Wege vom republikanischen Revolutionär zum europäischen Patrioten.

## 2. Der europäische Patriot

Schon Karl Hillebrands äußerer Lebensweg führt ihn kreuz und quer durch Europa, nicht als ein heimatloser Flüchtling, von einem Exil ins andere, sondern als ein Deutscher, der Heimat in Frankreich, in England, in Italien als seinem zweiten, dritten und zuletzt vierten Vaterland sucht, und findet. Auch der innere Lebensweg führt Karl Hillebrand nicht einfach von einem Lande in das andere, aus einer Sprache und Kultur in die andere. Auf diesen verschiedenen Stationen seines Weges durch die "Zeiten, Völker und Menschen" wird er allererst zum Europäer, dem sich aus immer neuer Perspektive ein anderer Aspekt dessen erschließt, was er *europäische Cultur* nennt.

Hillebrand sieht diese europäische Kultur, diese Kultur Europas, dieses Europa der Kultur ebenso vor dem Hintergrund ihrer *ursprünglichen Einheit wie heutigen Vielfalt.* Auch wenn "dieses Ideal nie völlig erreicht" wurde, dürfen wir doch (so sagt er) "das mittelalterliche Europa als eine große Familie betrachten, die eine Zeitlang glaubte, sie könne für immer unter *einem* Dache wohnen und gemeinsam an dem großen Werke der Civilisation arbeiten. *Eine* Sprache – die lateinische, *ein* Glaube – der katholische, *ein* Gesetz – das römische, *ein* Souverän – der Kaiser – sollten die Oberherrschaft führen und allen Gliedern der Familie Schirm gewähren" (VII, S. 1). "Europa entwuchs diesem Stammhaus, so geräumig es gebaut schien". "An dem Tag, an welchem ein philosophischer Gedanke in nationaler Sprache ausgedrückt wurde, hatte jene Teilung Europas begonnen, aus welche sich während des 15. Jahrhunderts die nationalen Monarchien von England, Frankreich und Spanien, die italienische Renaissance und die Reformation in Deutschland entwickelten" (VII, S.2).

Ausdrücklich sagt Hillebrand: "Die Teilung nicht die Spaltung", weil er auch die in der Vielfalt der Nationen des "modernen Europa geleistete Arbeit in Wahrheit als eine einzige" begreift, wenn sie nun auch in "Teilen, bald von dem, bald von jenem verrichtet werden muß", wobei "die Fackel des geistigen Lebens" gleichsam vom Einen zum Andern übergeht, wie er mit einem alten Bilde sagt (VII, S. 2).

Auch in diesem *historischen Prozeß*, aus dem die neuere europäische

Kultur entsteht, steht Alles mit Allem in Wechselwirkung: "Wenn der englische Empirismus eine Reaktion auf die spanische Dogmatik, eine Reaktion gegen den italienischen Humanismus gewesen, so war der französische Rationalismus, der im folgenden Jahrhundert die Oberherrschaft führte, eine Fortsetzung der intellektuellen Strömung in England, keine Opposition gegen dieselbe" (VII, S. 14).

Hillebrand wird ein Leben lang nicht müde, nach diesen geschichtlichen Zusammenhängen der geistigen Entwicklung auch und gerade der *modernen europäischen Kultur* zu fragen, um nicht nur den eigentümlichen Beitrag herauszuarbeiten, den dabei die verschiedenen europäischen Nationen geleistet haben, sondern zuletzt auch durch alle Vielfalt hindurch das allen Gemeinsame herauszufinden; was davon "integrierende Bestandteile der geistigen Verfassung Europas" (VII, S. 18) sind, also das was wir heute gerne als "héritage culturel européen" oder "patrimoine culturel européen" umschreiben, von dem schon Hillebrand sagt: "Wir haben ein gemeinsames Kapital von Ideen, an welchen wir alle zehren, in denen wir leben, oft ohne uns dessen bewußt zu sein" (VII, S. 18).

Schon mit diesen im Zeitalter des Nationalismus und der Reaktion wahrhaft "unzeitgemäßen Betrachtungen" betritt Karl Hillebrand ein neues Feld der *Kulturgeschichte und Kulturvergleichung in europäischer Perspektive*, das seinen Arbeiten zur *Kultursoziologie und Kulturpsychologie des modernen Europa* in unserer veränderten Welt neue Aufmerksamkeit verheißt. Jedoch nicht nur diese neue Sicht, sondern die neue wissenschaftliche Haltung, mit der hier das *Laboratorium Europas* aus der Perspektive Kultur betrachtet wird, scheint uns heute des geschichtlichen Erinnerns wert. Sie macht Hillebrand noch heute in unserer Zeit und Welt, wo noch immer geschichtliche Betrachtung aus nationaler Perspektive vorherrscht, zu einer ganz und gar unzeitgemäßen und darum auch für uns beispielhaften Gestalt.

Wer als Betrachter, so heißt es bei ihm, einem "Gegenstande, wie es der *Beitrag einer Nation zu Europas gemeinsamen Gedankenschatze* ist, gerecht werden will, der wird wohl daran tun, *sich allen Parteigeistes, des nationalen wie des politischen und des religiösen zu entäußern*. Der Parteigeist hat seine rechte Stelle im praktischen Leben". "Aber wenn wir es versuchen, die Geschichte der Menschheit zu verstehen und ihre geheimnisvollen Bahnen zu erkunden, ... da wollen und sollen wir solche unliebsamen Unterscheidungen vergessen und *einander behandeln, als ob wir alle zu einer Nation, einer Partei, einem Glauben gehörten*" (VII, S. 24; Hervorhebungen von mir). Hüten wir uns, so fährt Hillebrand darum fort, "Menschen und Tatsachen und Ideen zu verdammen oder heilig zu sprechen, weil sie russischer oder italienischer Herkunft sind, eine katholische oder protestantische Aufschrift tragen, aus dem konservativen oder liberalen Lager kommen. Dies würde nichts anderes als Barbarei sein" (VII, S. 25).

Diese Barbarei sieht Hillebrand im Erziehungssystem seiner Zeit im

Vordringen, das "nur zu Deutschen erziehen will", oder gar "zu Kaufleuten, zu Professoren, zu Ingenieuren und Militärs". "Sie sollten aber auch zu Menschen erzogen werden" (VII, S. 73), so folgert er darum, nicht weniger für seine, wie noch für unsere Zeit gültig.

Dabei ist es für Karl Hillebrand, bis in seine Bruchstück gebliebene letzte Arbeit hinein, vor allem andern die Rückbesinnung auf die unsere *Epoche der Moderne begründenden und noch immer bewegenden politischen Ideen*, die verwirklichten wie die unverwirklichten, die dem Menschen im "Nachdenken" ein Bewußtsein des Zeitalters gibt, in dem er lebt. "Wie viel näher steht uns da doch das 18. Jahrhundert! Wir teilen nicht alle seine Ideen, aber wir verstehen sie doch; sie bewegen noch heute einen großen Teil der Geschichte machenden Menschheit; Montesquieus konstitutionelle, Rousseaus demokratische Theorien, Voltaires Deismus, Condillacs Sensualismus sind noch nicht tot und begraben. . .; sie begegnen uns auf Schritt und Tritt, ja sie verlegen uns oft recht unbequem den Weg" (VII, S. 146). Wir fahren fort: Wie viel näher steht auch uns heute als der "geschlossene Handelstaat" Fichtes, oder selbst Hegels Staat der "sittlichen Idee", Montesquiueus Theorie der Herrschaftskontrolle durch Gewaltenteilung oder Rousseaus Theorie der Herrschaftslegitimation durch Gesellschaftsvertrag, die beide in unserer heutigen Theorie einer nicht mehr absoluten, sondern *konstiutitonellen Demokratie* fortleben; ebenso Kants Theorie einer "bürgerlichen Gesellschaft" der Gleichheit der Freiheit, ja schon Rousseaus Theorie einer "wohlgeordneten Gesellschaft" der Gleichheit der Wohlfahrt, deren beiden Gedanken sich in unseren heutigen Konzeptionen eines *freiheitlichen Rechtsstaates und Sozialstaates* wiederfinden. Auch wir leben so noch immer aus politischen Ideen, die in der bürgerlichen Aufklärung des 18. Jahrhunderts erstmals "zum Vorschein" gelangt, "in Gedanken" gefaßt sind.

Daß diese uns heute im 20. Jahrhundert näher angehen als selbst das 19. Jahrhundert, hat seinen Grund darin, daß in diesen *epochalen Ideen* unverfälscht und unverkürzt die Zielsetzungen einer *Demokratisierung und Liberalisierung des Staates und der Gesellschaft* insgesamt umschrieben werden, die im ursprünglichen Gedanken dieser demokratischen Revolution angelegt sind, in Hinsicht auf die Demokratisierung und Liberalisierung auch der Gesellschaft aber erst jetzt, in einer zweiten Phase demokratischer Evolution von der Mitte des 20. Jahrhunderts an, ins allgemeine Bewußtsein gedrungen und so geschichtsmächtig geworden sind.

Daß uns Kants Idee einer *Weltbürgerlichen Gesellschaft* der "öffentlichen Staatssicherheit" auch zwischen den "großen Staatskörpern", in unserem Zeitalter regionaler und universaler Entwicklungen nach eben diesem Prinzip, heute ganz anders angehen, als die aus dem Zeitalter des Nationalismus stammende Idee Hegels einer *Weltgeschichte als Weltgericht*, zeigt beispielhaft den grundlegenden Wandel der Zeit und Welt, von damals zu heute.

In einer solchen Rückbesinnung auf die für unsere Zeit und Welt lebenskräftigen Wurzeln der politischen Ideen, aus denen auch noch wir denken und handeln, wird so recht sichtbar und einsichtig, daß wir in der Tat in unserem "Laboratorium Europa" auch hier und heute aus jenem "gemeinsamen Gedankenschatz" leben, von dem Hillebrand spricht.

Auch wo uns Karl Hillebrand heute in seinen politischen Ideen ferne gerückt ist, verhilft uns seine zeitgemäße Gestalt: als *unbeirrbarer liberaler Republikaner* doch zu einem besseren Verstehen seiner Zeit und Welt. Wo er uns aber, nach einhundert Jahren, als eine damals unzeitgemäße, heute dagegen beispielhafte Gestalt unserer eigenen Zeit und Welt erscheint: als *verfrühter europäischer Patriot*, können wir uns selbst durch ihn besser verstehen lernen.

Wenn uns etwa Karl Hillebrand, um dieses ehrende Gedenken zum einhundersten Todestag mit seinen eigenen Worten zu schließen, für unser *Zeitalter der Parteiendemokratie*, nicht minder zeitgemäß als damals für seine Zeit und Welt des Nationalstaates, die sich im Irrsinn zweier Bruderkriege der Völker Europas selbst widerlegt, eine Mahnung vor der *Herrschaft des Parteigeistes* und der *Knechtschaft der Parteibande* zuruft, die noch immer, eher mehr noch Geltung und Gültigkeit hat: "Je größer die Zahl derer wird, welche am politischen Leben teilnehmen, desto mehr wird die politische, religiöse, nationale Leidenschaft der Gerechtigkeit, Billigkeit und Gutmütigkeit den Garaus machen. Denn ein jeder, der sich in die Knechtschaft der Parteibande begibt, muß notwendigerweise einen Teil der Wahrheit, die er kennt, einen Teil seiner moralischen und intellektuellen Freiheit, einen Teil seiner selbst opfern. Auf der anderen Seite wird bei denen, welche sich von solchen Leidenschaften frei machen, um die Dinge mit eigenen Augen zu sehen, nach eigenem Sinn zu beurteilen, die Liebe zur Wahrheit in demselben Maße an Kraft zunehmen, als ihre Zahl gering ist. Geben wir uns wenigstens die Mühe, zu diesen wenigen zu gehören; denn sie sind nicht allein die einzigen Freunde der Wahrheit, sie sind nicht allein die einzig freien Geister, sie allein sind auch die wirklich Gerechten" (VII, S. 25). Karl Hillebrand war ein solcher *freier Geist*, den sein Aufstehen für die Freiheit als *republikanischer Revolutionär* aus der angestammten Heimat vertrieben hat. Karl Hillebrand war ein solcher *wirklich Gerechter*, den sein Eintreten für die Gerechtigkeit, auch in vorbehaltslosem Wahrnehmen anderer "Zeiten, Völker und Menschen", zum *europäischen Patrioten* gemacht hat.

## Anmerkungen

[1] Vgl. dazu einführend: Wolfram Mauser, Karl Hillebrand, Band I, 1960, Gesetz und Wandel, Innsbrucker Literarhistorische Arbeiten, herausgegeben von Karl Kurt Klein und Eugen Thurnher.

[2] Karl Hillebrand, Zeiten, Völker und Menschen, Berlin, 1829-1884, zitiert nach der Ausgabe Karl J. Trübner, Straßburg, VII Bände 1885-1907.

[3] Vgl. zum folgenden im einzelnen: Werner Maihofer, Prinzipien freiheitlicher Demokratie, in: Handbuch des Verfassungsrechts, herausgegeben von E. Benda, W. Maihofer, H.J. Vogel, 1983, S. 173ff.

[4] Vgl. dazu im einzelnen: W. Maihofer, a.a.O., S. 205ff.

[5] Vgl. dazu im einzelnen: W. Maihofer, a.a.O., S. 212ff.

## The Contributors

JEAN BLONDEL, born in 1929, Professor at the University of Essex, has been a member of the Department of Political and Social Sciences since 1985.

GISELA BOCK, born in 1942, Professor at the Technische Universität Berlin, has been a member of the Department of History and Civilization since 1985.

IAN BUDGE, born in 1936, Professor at the University of Essex, was a member of the Department of Political and Social Sciences from 1982 to 1985.

MAURO CAPPELLETTI, born in 1927, Professor at the University of Florence and at Stanford Law School, California, has been a member of the Law Department since 1976.

EMIL-MARIA CLAASSEN, born in 1936, Professor at the University of Paris IX-Dauphine and at the European Institute of Business Administration (Fontainebleau) has been a member of the Economics Department since 1984.

HANS DAALDER, born in 1928, Professor at the University of Leiden, was a member of the Department of Political and Social Sciences from 1976 to 1979.

TERENCE C. DAINTITH, born in 1942, was Professor of Public Law at the University of Dundee, Scotland; he has been a member of the Law Department since 1981.

BRUNO DE WITTE, born in 1955, has been a Professor in the Law Department since 1985.

JEAN-PAUL FITOUSSI, born in 1942, Director of the research department of the Observatoire Français des Conjounctures Economiques and Professor at the Institut d'études politiques de Paris, was a member of the Economics Department from 1979 to 1984.

WOLFGANG GEBAUER, born in 1942, Professor at the University of Frankfurt, was a member of the Economics Department from 1983 to 1986.

RICHARD D. GRIFFITHS, born in 1948, is Professor of History at the Free University of Amsterdam.

PETER HERTNER, born in 1942, Professor at the Technische Hochschule Darmstadt, has been a member of the Department of History and Civilization since 1981.

WERNER MAIHOFER, born in 1918, Professor of Law at the University of Bielefeld, President of the Institute since 1981.

BERNARD MARIN, born in 1948, was Deputy Director of the Institute for Conflict Research in Vienna. He has been a member of the Department of Political and Social Sciences since 1984.

ALAN STEELE MILWARD, born in 1935, Professor at the London School of Economics and Political Sciences, was a member of the Department of History and Civilization from 1983 to 1986.

ATHANASIOS MOULAKIS, born in 1945, Professor at the University of California in San Diego, was a member of the Department of Political and Social Sciences from 1979 to 1986.

BIRGITTA NEDELMANN, born in 1941, Professor at the University of Freiburg im Breisgau, has been a member of the Department of Political and Social Sciences since 1985.

DOMENICO MARIO NUTI, born in 1937, Professor at the University of Birmingham, has been a member of the Economics Department since 1982.

JEAN G. PADIOLEAU, born in 1944, Professor at the Ecole Superieure des Sciences Economiques et Commerciales, was a member of the Department of Political and Social Sciences from 1983 to 1986.

JACQUES PELKMANS, born in 1946, Professor at the European Institute for Public Administration, Maastricht, was a member of the Economics Department from 1979 to 1982.

DANIEL ROCHE, born in 1935, Professor at the University of Paris I-Sorbonne, has been a member of the Department of History and Civilization since 1985.

PHILIPPE C. SCHMITTER, born in 1936, Professor at Stanford University, was a member of the Department of Political and Social Sciences from 1982 to 1986.

MANFRED E. STREIT, born in 1939, Professor at the University of Mannheim, was a member of the Economics Department from 1980 to 1983.

DANIEL SZPIRO is Director of Studies at the OFCE, Paris.

GUNTHER TEUBNER, born in 1944, Professor at the University of Bremen, has been a member of the Law Department since 1981.

KAMARASWAMY VELUPILAI, born in 1947, Professor at the University of Alborg, was a member of the Economics Department from 1982 to 1985.

JOSEPH WEILER, born in 1951, Professor at the University of Michigan, Ann Arbor, was a member of the Law Department from 1979 to 1985.

STUART J. WOOLF, born in 1936, Professor at the University of Essex, has been a member of the Department of History and Civilization since 1983.

UU0086006ENC