

9

Genesis and Structure of Evaluation Efforts in Comparative Perspective

Hans-Ulrich Derlien

Summarizing and comparing the various national reports presented in this volume is not an easy task to accomplish. The national development of evaluation depends on a number of factors and constellations. This context has historically grown and constitutes a part of national (and increasingly international) collective learning.

The conditions under which the evaluation function in the various countries developed are not merely interdependent, but they also influence the subsequent structure and operation of the evaluation systems that emerge. Nevertheless, for analytical purposes the complicated fabric first has to be unwoven into a summary which deals with the genesis, the historical process and sociopolitical roots of evaluation efforts. Part two will systematically treat individual change agents who played a role in the process of institutionalizing the evaluation function. Part three of the chapter will describe the elements which make up the present national evaluation systems (if there are such) and will try to outline the modes of their operation.

I should like to emphasize that this summary as a matter of course draws most of its information from the national reports included in this volume. Furthermore, I have tried to outline arguments presented during the meetings of the working group. To arrive at some coherent conceptualization, interpolations were necessary which were drawn from the relevant literature; thanks to the response of the working group members to the first draft I could eliminate most of the misconclusions and was in a position to fill in more accurate data.

Genesis of Evaluation Efforts

The overall impression one might derive from the reports is that policy evaluation has reached different developmental stages and degrees of maturity. This is so, both in terms of length of experience with this analytical tool to improve public policy by systematic learning from past performance of programs, and in terms of frequency, regularity, and ultimately degree of institutionalization of the evaluation function.

At present I would tentatively classify the state of the art in the countries surveyed here as follows:

- Undoubtedly the most advanced evaluation system can be observed in the United States. Looking back on roughly twenty years of experience with policy evaluation since the 1960s, the American system has reached a high degree of methodological sophistication. This development allows one to clearly distinguish PE from other more traditional techniques of accounting and auditing. In addition, the U.S. evaluation system is firmly institutionalized within the bureaucracy and within the legislative branch (including the General Accounting Office (GAO)) and consequently produces a huge number of studies. Furthermore, the American system seems to have affected the philosophy, the skills, and the willingness of other countries to introduce policy evaluation, too.
- With the first examples of policy evaluation reaching back into the 1960s, Sweden, Canada, and the Federal Republic of Germany (FRG) may be distinguished as a second group of countries following the U.S. efforts relatively early. However, despite institutionalization of PE in various policy areas, their systems remained rather fragmented and the number of studies carried out seems to be relatively low.
- Great Britain and Canada both, apparently, have reemphasized PE at the end of the 1970s after earlier attempts at planning and evaluation had been regarded as unsuccessful. The two countries, however, have institutionalized the evaluation function in different ways. While Canada attributed evaluation to a Comptroller General as well as to individual government departments, in Great Britain evaluation hardly seems to have an organizational locus, but rather appears to be part of the general management philosophy.
- It would also be possible to put Great Britain together with Norway, Denmark, France, Switzerland, and the Netherlands—those countries which in the 1980s have emphasized policy evaluation beyond ad hoc studies in an effort to increase productivity in government. Some of them have linked policy evaluation closely to the budget process.

In the following sections, the history of the various national evaluation efforts is to be more closely inspected, and some contextual factors are

elaborated which may well have influenced the propensity of the various national systems to move towards policy evaluation.

Historical Account of Institutionalization

The roots of the U.S. evaluation efforts traditionally are seen in President Johnson's War on Poverty programs of the mid-1960s and the concomitant efforts to rationalize policy making by institutionalizing the famous planning programming budgeting system (PPBS) in 1965. It is worthwhile keeping in mind this connection between intervention policy and PPBS on the one hand and evaluation on the other. While the federal departments, above all the Office of Economic Opportunity, created evaluation staffs to monitor the results of their welfare programs, the U.S. Congress, not wanting to rely exclusively on executive branch studies, required in the Economic Opportunity Act of 1967 the General Accounting Office to assess the effectiveness of the poverty programs. Under the conservative Nixon administration, attempts were made to strengthen the evaluation function at the federal level by assigning to the Bureau of the Budget (BOB)¹ the task of stimulating departmental evaluation studies and fostering the spread of methodological know-how. Since then both these branches of U.S. government, the executive branch and Congress, have continuously increased their efforts to make evaluation a regular procedure. Despite the 1979 Office of Management and Budget (OMB) circular A-117 renewing the request for evaluations, legislative efforts seem to have contributed more to the persistence of PE in the U.S. system, be it the introduction of Sunset Legislation (Adams and Sherman 1978), or the institutionalization of the Institute for Program Evaluation in the GAO in 1980, with nearly one hundred professional staff (later renamed Program Evaluation and Methodology Division). These capacities, along with other congressional analytic capacities, are the more important in recent years, as the executive branch departments have dramatically reduced their evaluation units since 1980.

In the FRG, too, the emergence of PE is closely connected with intensifying social and economic interventions and reform policies under the government led by a Social-Democrat chancellor in 1969. The landscape, however, is much more fragmented, and evaluation has not been as firmly institutionalized as in the United States. Some departments have created evaluation sections, while others have their programs evaluated under the cover of general research activities commissioned to outside contractors. Most characteristic for the early 1970s was a number of experiments (basically in education) which, however, did not reach the level of meth-

odological sophistication of the large-scale U.S. experiments (Rossi and Wright 1977).

Apart from these program-specific institutionalizations, PE became institutionalized in the budget process in 1973 and thus enables the Finance Ministry to request studies from individual departments. This, however, has seldom been done. An equally weak position has a methods section in the Chancellor's Office, which does not even stimulate the departments to carry out evaluations, nor does it assist them in their evaluation attempts.

Although the necessity to evaluate programs has been emphasized as a logical consequence of the planning cycle and the concept is generally accepted, an evaluation culture has not yet developed. This may be due to the somewhat ambivalent role the German federal Parliament plays in this regard: on the one hand there are occasional requests for evaluation studies and legislated evaluation demands; on the other hand one has to state a lack of discussion of the roughly two hundred annual reports (including evaluations) which Parliament receives. The scientific staff of the Bundestag is too small and lacks policy analysts. Neither has the German Federal Audit Office played an active role in enforcing an evaluation system on the executive branch, nor has it carried out systematic studies itself; recently, however, evaluation has been acknowledged as a task of the GAO in principle.

Like the German government, the socialist Swedish government, too, was preoccupied with launching and expanding programs until the mid-1970s, while relying for oversight on traditional accounting and financial reviews. Although recognized in theory as a necessary part of the new planning, programming, and budgeting-influenced administrative systems, policy evaluation was much less important than planning, programming, and budgeting. In contrast to the FRG, the reforms relied on a broad political consensus typical for Sweden, and the programs were based on extensive *ex ante* analysis in committees including social scientists. PPBS, as a concept, has had a greater bearing in Sweden than in the FRG. Not surprisingly, then, effectiveness auditing and program-cost accounting were introduced as early as the mid-1960s. In the central government this development was inspired and implemented by the National Audit Bureau (Ysander 1983, 10). A great portion of its audits are policy evaluation carried out by some one hundred academics trained in economics, social science, and business administration. In view of economic stagnation a period of broad reappraisal of existing policies started at the end of the 1970s. Both Social-Democrat and nonSocialist (1976–82) governments have emphasized the need for more thorough appraisal of policies and programs. Evaluation units have been established within

supervising national agencies. In addition, expert groups attached to ministries have redirected some of their interest from planning to evaluation. Still, however, policy evaluation cannot be considered as institutionalized in the ministerial work or in Parliament.

The reappraisal of policies rather than of programs seems to have been also the task of the British Program Analysis and Review (PAR), coordinated by the Central Policy Review Staff (CPRS, 1970), and the Treasury. This central PAR, however, faded by the mid-1970s without recognizably putting its stamp on departmental-review activities. Despite formally abolishing PAR in 1979 and CPRS in 1983, it was the conservative Thatcher government that reactivated the early evaluation impulse in its emphasis on "value for money" and in its managerialist policy approach. While the 1983 Financial Management Initiative concentrated on inputs, nevertheless, prospectively new policies were required to make provision for subsequent evaluation. Further, in 1985 a Joint Management Unit was created with the aim of making evaluation a regular activity of the departments.

The recent efforts of the British government have been accompanied by increased parliamentary emphasis on evaluation and by strengthening the Public Accounts Committee (PAC). For example, the National Audit Office established in 1983 was not only to replace the Exchequer and Audit Department, but also to serve the PAC. It remains to be seen if this new system really arrives at evaluation or merely effectuates traditional accounting procedures.

Historically and systematically Canada probably comes closest to the constellation under which policy evaluation became institutionalized in the United States. Not only did Canada adopt the PPBS in 1969 (and retained it), but also the Treasury Board set up a Planning Branch, which developed evaluation capacity and support for departmental evaluation efforts. Nevertheless, by the 1980s policy evaluation turned out to have been largely unsuccessful in Canadian central government. Again, it was the legislative branch, in this instance the Canadian Parliament, which in 1977 with the help of the Auditor General pressed the government to create the Office of Comptroller General. This office contained a special Program Evaluation Branch to promote policy evaluation as part of the management cycle.

An interesting Canadian feature is the resulting impetus of these parliamentary initiatives to institutionalize policy evaluation with the permanent deputy head of departments and to create evaluation units in each department. At the same time policy evaluation within the Policy and Expenditure Management System (introduced in 1979) is related to the executive budget process through the Treasury Board's requirements to

submit evaluation results and evaluation plans. It is indicative of the degree of institutionalization that in Canada (as well as in the United States) government manuals exist, which prescribe the formal properties and the methodology of evaluation studies.

While the United States, Sweden, and the FRG have continued their early evaluation efforts into the 1980s (although with different degrees of institutionalization and frequency of studies), and while Canada and Great Britain have taken a second attempt in the late 1970s, there is a third group of countries where the use of evaluation studies was limited (Netherlands) or almost absent (Denmark, Norway) until quite recently.

In the Netherlands a government reform commission in 1980 and, more recently, Parliament pressed for evaluation as a regular policy-making device, after earlier attempts by the Finance Ministry to promote programming and policy analysis had failed. The Center-Liberal government in 1981 took initiative to procedurally institutionalize policy evaluation with the annual budget process (Reconsiderations Procedure). As in Britain, the driving force behind policy evaluation is the government's aim to roll back the frontier of the state and to reverse a long political tradition of social-intervention policy.

The philosophical link to Britain is even more obvious in the Norwegian case, where evaluation was institutionalized in 1985 as part of the reforms in the government budgetary system, following a productivity campaign in 1982. Furthermore, in Norway the auditor general, who reports exclusively to the legislature, has adopted performance auditing in addition to traditional financial auditing—almost twenty years later than Sweden, whose administrative innovations are normally readily introduced in Norway.

Finally, in some countries, such as Denmark, the concept of evaluation has spread in the 1980s. Evaluation studies are mainly undertaken by research institutes, although the device is not formally institutionalized in national government. In addition, since 1973 the Danish Department of Administration and Reorganization, together with the Ministry of Finance, has regularly evaluated various policy areas. France serves as an example of a country which relies heavily on traditional legal and accounting control instruments. Only in 1982 (under the Socialist presidency) was evaluation stressed as a necessity. In 1983, a survey of the audit and research reports produced so far was undertaken in order to find out the extent to which they have evaluative contents (Quermonne and Rouban 1986). Switzerland at best seems to feel the need for PE at present; this case will serve us for discussing factors which could have inhibited the development of policy evaluation.

This rough outline of the development of PE in the various countries reveals three trends:

- In the 1980s there emerges a second movement concerned with questions of output and, to a lesser extent, impact. Even most of those countries that institutionalized evaluation at the end of the 1960s experienced a second shift of evaluation activities.
- While the first evaluation movement of the 1960s and 1970s was closely linked to the planning and programming process and thus to the role of the program administrator, the attempts of the late 1970s and the 1980s are, rather, geared to either the political level, reconsidering the justification of policies, or to the budgetary process. In both cases the role of the external auditors became important.
- The locus of evaluation consequently has shifted in the 1980s from being primarily an internal government operation to becoming a concern of parliaments. Concomitantly, the actors have changed, too. Now one finds the Parliament as a key center of activity along with those auditing offices that primarily report to Parliament and have, therefore, been brought into the game.

We may ask now, which factors determined the variation in maturity of the evaluation systems and which constellations were influential in the observed individual developmental patterns. In the following section, I shall discuss exogenous factors responsible for the state of the art, as well as those endogenous factors, that is, characteristics of the internal environment of national government, which can be isolated as change agents.

Fiscal Situation

The distinct developmental patterns of the early starters—the United States, Sweden, Canada, and the FRG—from most of the other countries following along at the end of the 1970s or later, might be related among other factors to the conditions of the national economies and thus to the fiscal situation of the respective governments.

The first group of countries got the impulse for PE during the 1960s, that is, a time of booming economies and growing budgets, which enabled governments to embark on expensive social intervention programs including education and health care. In this context formal planning systems emerged, which either were limited to medium-term financial planning (FRG) or even attempted to integrate budgeting with programming (the United States, Sweden, Canada). In any case, evaluation was either regarded logically as part of these planning systems or as necessitated by the information needs of the intervention programs, the results

of which could hardly be predicted with a sufficient degree of certainty. Evaluations, then, were primarily used by program managers to effectuate existing and new programs. Only occasionally were evaluations necessary to protect these programs against political opponents by proving their effectiveness.

The impact of the totally different economic situation of the late 1970s on evaluation was almost the same: shrinking economies and the necessity to manage scarcity induced stress on governments and led to the development of cutback management techniques (Schick 1988). The way by which the relatively unfavorable economic and the critical fiscal situation promoted evaluation was via the budget process: financial management initiatives (Great Britain, Canada, Norway, the Netherlands) either made use of previous evaluation efforts (Canada, Great Britain) and thus engendered a second wave of evaluation efforts, or the new managerialism introduced the concept of evaluation in order to rationalize resource allocation within the budget (Norway, the Netherlands). Typically, the bearers of this second evaluation movement are not the program administrators in the government departments, but rather the finance ministers and the auditing offices as the traditional wardens of the budget. Consequently, the perspective on and the function of evaluation slightly shifted: instead of effectuating programs, the emphasis was rather on curbing ineffective programs in order to cut back the national budget.

Political Constellation

Of course, it is not the economic situation per se which brings about changes in policy direction, but ultimately politicians and government parties change the course of the ship of state. In the booming 1960s the countries of the first evaluation wave were governed by reformist parties (Social-Democrats in the FRG and Sweden, Democrats in the United States, Liberals in Canada) who engaged in the various reform policies and had a clear affinity to the employment of social-science methods needed for policy evaluation. Another case in point might be the Socialist presidency in France (1981), which put an emphasis on evaluation in 1982.

Rather conservative governments (the United States 1968–74, 1979–present; Great Britain 1970–1974, 1979–present; Canada 1979, 1984–present; Norway 1981–86; Denmark since 1982; Sweden 1976–82; Netherlands 1977–81, 1982–present), often facing a less prosperous economic situation, made use of existing evaluation instruments not only to re-tailor the budget, but occasionally in an attempt to curb reform programs of their predecessor governments, which they had viewed with suspicion

from the beginning. This is not to mean that interventionist political parties, were they in a position to govern, would not have had to respond to the changed economic situation in a similar way. In addition, conservative governments tend to look at social scientists with suspicion or to feel that they are the advocates of the former reform policies; their enlightening function is regarded skeptically. On the other hand, where analytical capacities were firmly established, incoming conservative administrations could try to use them as means of "analytical delegitimation" (Wagner and Wollmann 1986a).

Not surprisingly, there is also a neoconservative international consensus that the state should withdraw from society or should at least not further expand, and that many a program had overshot its goal. In connection with policies of privatization, deregulation and de-bureaucratization, scrapping the planning system, which had served to develop reform programs, was one device to change the course (United States 1970; FRG 1982; Great Britain 1979; Denmark 1983). The other was to improve the traditional budgetary track by importing accounting techniques from the private sector — techniques that had to deal with output (less with impact) of programs as the unit of analysis and that therefore resembled evaluation methods. After all, modernized accounting techniques fitted the traditional auditing procedures of the audit offices.

Constitutional Features

An explanation of the distinctive developmental paths followed by the various national systems that concentrated solely on the correlation between expanding economy/intervention programs/social-democrat governments or shrinking economy/austerity policy/conservative governments would be too simple a model (Wagner and Wollmann 1986a). It would, for instance, not explain the change in the United States starting under the Republican Nixon administration as early as 1970, nor would it explain why almost nothing changed after a conservative government had taken power in Bonn in 1982. Of prime importance seems to be the constitutional relationship between the executive and the legislative branches of government as an additional variable.

In parliamentary party government systems the classical theory of division of power hardly holds true empirically; here the cabinet has a strong hold on the parliamentary majority or shares the interests and programmatic consensus of the majority faction(s), while it is merely the opposition faction(s) who have a strong motivation to control the executive, but who regularly are not in a position to enforce controls on it. The short historical account presented earlier has shed light on the propagating role

of parliaments in systems of party government (Canada, Norway, Britain). It might, however, be supposed that parliaments merely emphasized what the administration wanted anyway. The German experience of an active parliamentary role with respect to the institutionalization of evaluation in 1968–1969 is a case in point: during this period a grand coalition was in government. In general, however, there are no visible differences between the political parties as far as policy evaluation is concerned.

The situation might be completely different in presidential government systems like those of the United States and France, where parliamentary majorities may form against a president and enforce measures on an administration in order to enhance their control capacity or informational independence. What wonder that a powerful U.S. Congress built up a huge congressional staff and made use of this counterbureaucracy to check the program administrators in the executive branch? In such a setting the legislative branch becomes a strong pillar in the evaluation system even when evaluation efforts originated in the administration. Legislative interest might help to keep executive evaluation capacities in place, even if the administration is not too enthusiastic about it, because the administration then needs these analytic staffs to keep the informational balance of power vis-à-vis Congress. France deviates from this configuration, because Parliament is constitutionally weaker and the executive branch is rather self-sufficient in its orientation towards using social-science policy advice in general (Wagner and Wollmann, 1986b).

Another constitutional factor which could be of importance in this respect is the affiliation of the central accounting or auditing unit. In most countries these offices are regarded as assisting units of the legislative branch and report to parliament (United States, Canada, Norway, the Netherlands, Denmark). The GAO in the FRG has historically been attached to the executive branch, and only since 1969 has it slightly redirected its function also to serve Parliament and the small parliamentary audit committee. Similarly, the French *Cours des Comptes*, although reporting to Parliament, too, has an executive bias owing to the career patterns of its members (Nioche and Poinard 1985). In Sweden with its tradition of small ministries and highly independent agencies, the dominant audit institution, the National Audit Bureau, is exclusively an instrument of the central government. Its reports are either directed to the audited agency or to the government when principle matters are brought up. The parliamentary auditors here have a small staff to undertake ad hoc scrutinies.

The more a central auditing unit is independent of the executive branch and even attached to parliament, the more parliament may be tempted to use these institutions as a power basis and a source of counteradvice vis-

à-vis even a party government. In the American case this arrangement, of course, could but strengthen congressional evaluation efforts.

Supply of Social-Science Knowledge

Evaluation methods and policy analysis in general have nowhere been invented by the bureaucracy, but have been developing in the realm of the social sciences and economics. There are strong indications that the developmental stage of the respective government evaluation culture is decisively dependent on the propensity of the academic world to produce evaluation devices, engage in evaluations, and professionalize the roles of evaluators and policy analysts in general (Wagner and Wollmann 1986b).

Undoubtedly, the leading role of the United States in institutionalizing evaluation procedures and in creating evaluation staffs in government to a great deal also reflects the long tradition of applied (quantitative) social research. This supply seems not to be readily available in most of the other countries. In these countries there is either no tradition of applied social-science research in universities, or the professionalization of the social sciences (curricula, scientific associations, specific jobs) has only started in the 1970s.

These diverging antecedent conditions also could have indirectly affected the receptivity of the administration for the diffusion of evaluation skills, as they determine the probability that social scientists or economists are recruited at least into bureaucratic staff positions. This was the case in the United States and Canada and, to a lesser extent, in the FRG and Great Britain (Bulmer 1988). The Swedish case is indicative, too: in the 1960s economics, business, and public administration graduates reached leading positions in the public sector and were also the only recruits of the National Audit Bureau, when it started effectiveness auditing.

If the bureaucracy (and politicians in the legislative branch) have overwhelmingly undergone a training in law (France, Germany, Denmark) or even classics (Great Britain), the system might be less receptive for social scientists and social-science knowledge.

Furthermore the existence of independent, non-university research institutes providing evaluation capacities seems to depend on a strong tradition of applied social sciences. In a number of countries new policy-research institutes were founded in the 1970s (see the subsection "Internal versus External Evaluation Research" in this chapter).

Clearly, these factors explain differences in the supply of evaluation studies, and by way of staff affinity to the social sciences also affect the demand for this kind of information. A lack of scientific infrastructure

would also aggravate the institutionalization of policy evaluation even when the need for this kind of analysis arises.

Functional Equivalents

It is equally interesting, however, to ask which conditions — *ceteris paribus* — explain the lack of demand for evaluation research. Among other factors, it might be the availability of functionally equivalent kinds of information, in particular good official statistics. One could possibly maintain that in continental Europe and Scandinavia, that is, in countries with a long (and still strong) state tradition, the statistical apparatus is one of the backbones of the bureaucracy on which policy analysts (and economists) customarily rely. A case in point could be the Netherlands with their long planning tradition originating in Tinbergen's efforts. On the other hand, historically it was the very absence of a reliable statistical apparatus which gave rise to applied social science in the United States.

Second, apart from the census system, internal administrative data (process-produced information) are possibly not evenly available in all of the countries under investigation.

Another equivalent, which for quite a long time seems to have satisfied the rising evaluative-information demand in all of the countries, is the existence of an advisory committee system around central government. This feature was reported for Sweden, where almost all reforms were developed and reassessed by ad hoc committees; for Denmark; and for the Netherlands. Policies are here assessed by experts from science, interest groups, and bureaucratic professions; this system could at least offer what project managers are not seldom most interested in: implementation assessments.

Last but not least, some evaluative information, which indicates the need for program amendment, is produced by the system of administrative courts in countries with a strong public-law tradition or by the ombudsman (Denmark, Sweden, Norway, the Netherlands). Investigative journalism seems to play an important role in countries that disclose documents under freedom of information acts (the United States, Sweden, Canada).

Retarding Factors

Of course, all of the countries under scrutiny have accepted by now the need and the notion of policy evaluation. In trying to sort out factors that could explain the different developmental stages the various countries

have reached, we should also mention some of the factors which possibly contributed to a slow or late development of policy evaluation.

The above-mentioned functional equivalents surely might have retarded the introduction of policy evaluation. As far as features of the old continental state apparatus were concerned (official statistics, administrative courts), these point to the public-law tradition as a more general condition. Recruitment of jurists into the political (to a lesser extent) and bureaucratic (France, Germany, Denmark) elite shapes the role understanding of actors in all subsystems, parliament, administration, and auditing offices in a peculiar way: questions relating to efficiency and effectiveness are at best asked in the second instance; to think in terms of programs (with stated goals and projects to achieve them) is rather alien to people who are used to consider laws in principle as made for eternity, although practically laws are frequently amended and increasingly aim at specific effects to be achieved. Furthermore, the external control apparatus (administrative courts, auditing offices) primarily investigates the legality of decisions and actions, thus moving the focus away from the effectiveness of programs. In such a politico-administrative culture, the idea of policy evaluation is more likely to find reception and propagation in the realms of social science and economics, and its extension depends on these professions finding their way into core institutions.

We also considered whether the federal or unitary constitution of a state had an effect on the extent to which PE was applied. Have the centers of federations a higher need to monitor implementation? It would be plausible to assume such an information need where programs are executed by offices of member states (FRG, Canada), but less where the central government executes programs through its own field offices (United States). Thus, in Canada and the FRG, where joint federal-state programs on a shared-cost basis were launched, evaluations were institutionalized to make sure that programs met national standards. It is indicative that the Swiss federal government in 1987 institutionalized a think tank to explore evaluation possibilities.

Monitoring is more easily accomplished in unitary states where the center can request the field offices to provide statistics. In a federal system with administrative autonomy of the members, however, it is occasionally not feasible, at least politically, to demand detailed information from the states, unless the procurement is especially institutionalized.

Another variable of potential importance could be the extent to which government programs and services are delivered by nongovernmental organizations, be it by commercial organizations or by commonweal, non-profit organizations. Here in principle the argument raised above (federalism) would hold true as well. A new constellation, though, emerges

when services, which elsewhere derive from government programs, are provided completely privately, that is, under societal self-regulation and self-help. Obviously, here the volume of state intervention and activity would be rather small, at least in particular policy areas, and the need to monitor effects of government policies would be minor (although in rationalist perspective the demand for social indicators and societal monitoring would increase). This argument can be related to federalism again: of course, central government's information needs decisively depend on the division of jurisdictions between center and subsystems; where, as in Switzerland, the federal government has centralized only few competences, we would not expect much policy evaluation. Would the need for PE then be felt the stronger on the subgovernmental level?

This line of reasoning leads us to another factor of potential importance: the size of a country and administrative distance. Smaller countries (in terms of territory, population, or both) might have smaller problems with less local variation and less need for central intervention in many policy areas. This hypothesis, however, does not hold true in the cases of Denmark and Sweden with their large public sectors. If in addition to small size the political system is decentralized (Switzerland), or the administration is de-concentrated onto the local level (Denmark), the administrative distance between the population/problems and the regulating authority is small, and feedback from society reaches the steering bodies and political institutions more directly. In the most extreme case where there are few general societal problems and many locally confined problem areas dealt with by autonomous local government units, the impact of measures could be directly inspected, cause and effect could easily be related to one another, and PE as an artificial feedback mechanism would be regarded as superfluous. However, with central government programs operating in such a decentralized structure, the need for central evaluation seems to increase, if uniformity with national goals is regarded as necessary.

Obviously, it is hardly possible to attribute the developmental stage of PE exclusively to one of the factors mentioned above. After all, the search for an explanation of the observed differences should not distract our attention from a very common phenomenon: changes in the repertoire of tools for managing public affairs are heavily subject to fashions and fads. Administrative philosophies (in particular planning systems and management styles) come and go like tides. New ideas are often introduced from private management (Sweden in the 1960s) and are then transferred from one government to the other, not merely because they have proven useful, but also because governments want to be thought of as modern and to view themselves as innovative. Cross-fertilization was

reported for Scandinavia, where Sweden plays a leading role, and for North America, with the United States having provided the example for planning and evaluation. In general, the spillover of concepts is facilitated where countries formally cooperate and specialists gather in international associations like, for instance, the supreme audit institutions (INTOSAI).

Once a reform has been carried through, the new management structures press for consistent control mechanisms. A case in point can be observed not only in the evolution of evaluation from formal planning systems, but also after the introduction of management by objectives and decentralized budget spending in Sweden: both measures increased the need for evaluative information to keep the management system operating.

It might be justified to summarize the historical process leading to the institutionalization of PE in a two-stage model. While a first wave of evaluations in the 1960s was closely linked to social-reform policies of social-liberal governments operating under favorable fiscal conditions, a second wave of the evaluation movement was stimulated by predominantly, although not exclusively, conservative governments' attempts to curb intervention programs in view of fiscal strains. Consequently, there is a tendency to reorient PE from its predominant focus on programs and their improvement towards the budget and its curtailment. A concomitant shift of the central actors can be observed, too: instead of departmental program managers, finance and audit institutions, in particular those serving functions for national parliament, become protagonists of policy evaluation.

This picture by no means holds true for all countries surveyed. There are quite a number of continental European countries where a variety of constitutional, administrative, and cultural factors have hampered the process of institutionalizing PE as a regular activity in the politico-administrative system. Nevertheless, even in these countries the demand for (ad hoc) evaluations seems to have increased in response to the general political and economic forces operating in the early 1980s.

Change Agents

Once the demand for evaluative information is perceived and acknowledged in a polity, it is nevertheless still far from being satisfied. It is almost an anthropological proposition that people do not like to be supervised and resist having their activities monitored, because they basically fear negative sanctions in general and threats to their careers in particular resulting from the documentation of failures. In addition,

external controls—by definition—change the existing balance of power between organizational systems and therefore are resisted for political reasons. It is also frequently the case that resistance arises not only when new control measures are being introduced, but is already arising when concepts are discussed. Therefore it is worthwhile to look at the agents which brought about change, even in the face of resistance from other parts of the national polity.

Parliament

National parliaments have been pressing their executive branches to introduce policy evaluation. In the FRG the legislative branch demanded evaluation studies in the fields of developmental aid and technology policy in the late 1960s, but did not specialize internally for this function, nor did the Parliament establish specialized staff capacity. In Canada, Parliament took up the initiative of the auditor general in 1977 and entitled him to supervise executive evaluation efforts. In Britain, too, Parliament since 1979 strengthened the system of select committees. These initiatives always addressed the executive to produce evaluations. In the United States the role of Congress for constitutional reasons was more powerful: apart from legislating evaluation requirements in laws (as did parliaments in other countries, too) and introducing sunset legislation, Congress, with the help of the GAO and other staffs, built up its own evaluation capacities to conduct its oversight responsibilities vis-à-vis the executive branch.

In other countries (the Netherlands, Norway) parliaments have changed their budgeting procedures and put more emphasis on evaluation in full consent with governments.

Supreme Auditing Institutions

In most of the countries, in particular in those which followed on the second wave of the evaluation movement, the auditing offices played an active role. In Sweden performance auditing, of which policy evaluation is an integral part, became the core function of the National Audit Bureau (NAB) as early as 1967, while financial auditing (until 1987) had been moved to local offices. Furthermore, the NAB promotes evaluation by issuing guidelines and offering seminars. In the United States the GAO established its evaluation office in 1980; in Canada the auditor general pointed to a need to improve the policy evaluation function since 1976 and induced Parliament to institutionalize his role in this respect more explicitly. A similar development took place in Britain, where the Exchequer and Audit Department was replaced by the National Audit Office in 1983. In Denmark cost accounting was emphasized by the national audi-

tors in 1985. In the Dutch and German cases the GAOs have rather played a reactive role. Although they are entitled to review all government operations with respect to efficiency and effectiveness and to assess evaluation procurements in their executive branches, their traditional role of checking conformity with budget law and controlling formal correctness of operations, which they have in common with all accounting offices, seems to be still dominant. Nevertheless, their role understanding is changing in response to growing concerns about efficiency and effectiveness in other parts of the national policy. The Dutch GAO has produced a number of reports on large projects. The German Bundesrechnungshof in 1985 has finally explicitly accepted the evaluation function as part of its broad supervisory role and even carried out a survey of the state of the art in federal departments in 1986. To date this survey has not been published.

The Cabinet

Has evaluation been a concern of cabinets; that is, have there been central political initiatives in the executive branch? Either the findings in this respect are in inverse proportion to the rather active role parliaments and auditing offices have played, or cabinet initiatives were complementary. Some cabinets took broad political initiatives to increase productivity in general: for example, the Netherlands' Reconsiderations Procedure in 1981, Norway's 1982 Productivity Campaign, Britain's 1983 Financial Management Initiative, and Canada's 1979 Policy and Expenditure Management System. In Sweden, the government's leading policy has been to convince agencies to evaluate their programs, increase their effectiveness, facilitate adaptations to across-the-board cutbacks, or all three. These cases belong to the second wave stimulated by unfavorable economic developments and their governments' concern to manage the state budget more efficiently and to limit its further growth. That the early starters in the 1960s lacked central policy initiatives to establish evaluation capacities is plausible, because their administrations' concern was with developing reform policies. While evaluation in those days might have appeared to be a nonpolitical technical device, in the 1980s managing the budget has become an issue of political priority and therefore has attracted the cabinets' attention and paved the way for evaluation techniques.

Central Government Units

This is not to say that only recently would one find central units in national government dealing with evaluation. The U.S. OMB had taken on the role of advising the departments to conduct evaluation studies as early as 1970, and it reemphasized this concern in 1979. In Britain, too,

the Central Policy Review Staff within the PAR procedure had tried to motivate departments to undertake studies, as has Rayner's Efficiency Unit since 1979. In the Netherlands attempts had been made to strengthen policy analysis in general by establishing the interdepartmental Committee for the Development of Policy Analysis in the 1970s and, as its successor in 1985, the Agency for Policy Analysis under the purview of the Ministry of Finance. Canada is the outstanding example of a recent creation of a special central office concerned with evaluation: the Office of Comptroller General, created in 1978. Here the mandate of the office is to function as a change agent towards more evaluation in the departments.

One probably may generalize that, where central units exist, they took on the task of promoting evaluations throughout government, to develop guidelines and give assistance to departments. Their role is nowhere to evaluate departmental programs themselves. Obviously, it has been recognized in all of the countries that such an interference with (partly formal) departmental autonomy would have seriously driven up the level of conflict in (coalition) governments.

When regarding central government units one should also take into account the department or office which centrally controls the executive budget cycle. The organizational position of the U.S. OMB as an office of the White House is outstanding and has enabled it to exert authority in promoting evaluations. Apart from this authority derived from its formal position, the degree to which evaluation results are linked to next year's budget seems to be of prime importance for the success of this central solution. Normal finance ministries can acquire the same influence over departmental evaluation activities, once the latter are related to the budget process (and thus sanctioned and reinforced). For instance, the Dutch Ministry of Finance has its hands in the intradepartmental study groups which prepare reports for the Reconsiderations Procedure. One reason why the Finance Ministry in the FRG has hardly influenced the departments despite its formal right to request studies is the fact that it has seldom used this authority and, thus, never integrated evaluation with budgeting. The Danish rotating evaluation system performed by the Department of Administration and Reorganization annually selects critical issues from varying policy areas for evaluation, but tends to be a consultant to the departments rather than a budget controller. The Swedish Ministry of Finance established an independent expert group in 1981, which—inter alia—produces ad hoc evaluations in all policy areas, and the Dutch Ministry of Finance is assisted by an Agency for Policy Analysis.

It seems to be the advantage of the recent neoconservative manage-

rialist over the previous program orientation of evaluation to establish precisely this link between evaluation and budgeting. It remains to be seen, however, what effects this may have on the substance of evaluations. One potentially negative outcome could be a decrease in the level of methodological sophistication. Another could be that the budgeting process will be confronted with biased success stories and not objective data as justifications for additional funds.

Departments

It is hardly possible to generalize about the role the departments have played in bringing about policy evaluation. Clearly central political initiatives or external pressures were needed in many cases to move them towards evaluation. On the other hand, there have always been some departments that started evaluations on their own initiative. Take, for instance, the then U.S. Office of Economic Opportunity or, in many countries, those departments involved in developmental aid. In these instances it was the information requirements of the programs, a number of social scientists among the personnel, and a minister committed to one program and its improvement or international practice stimulated by the United Nations or the World Bank that led to the establishment of evaluation staffs and to the commissioning of evaluation studies. Again, the recent Canadian move to make the deputy ministers responsible for departmental evaluations is unusual and reflects the central political initiative behind it. The normal case is probably that departments simply have programs or projects investigated as part of their ongoing normal research activities and that, for such efforts, they do not even term these studies "evaluations".

Where the quantity of these studies grew, departments were inclined to institute evaluation staffs or methodology sections in general. Frequently behind such a development stood the external pressures by parliament or the finance minister to produce a report (of any quality whatsoever) or to submit a cost-benefit analysis (as in the FRG).

As to the policy areas in which PE came to be practiced relatively early and which are being carried out most often today, it is probably not an overgeneralization to name developmental aid, labor market policy, health, and education. These were also the most common target areas of reformist policies.

The predominance of decentralized evaluation activities in the early developmental years also poses a methodological problem for writing a report like this. It is difficult to discover a system and, unless by some chance there have been observers in the field, in retrospect these early

efforts may escape from our attention in view of more recent and spectacular central and systematic innovations. If longitudinal data on the number of accomplished studies were available, we could assess the decentralized activities more accurately; the availability of this type of statistics, however, would presuppose the very existence of a powerful evaluation movement.

Structure of Evaluation Arrangements

In this last section an attempt will be made to complement the genetic perspective by a systematic account of the evaluation arrangements devised in the various countries. The approach followed here will be procedurally oriented regarding evaluation as a process with distinct steps. The structural factors which have been analyzed above will be referred to, but will not constitute the core of the analysis. Thus, only those aspects of the activities of a specific institution, for instance the supreme auditor, which are relevant for understanding the evaluation process will be mentioned.

Initiation of Evaluations

Unless conducting evaluation studies becomes institutionalized, their occurrence tends to be random. Furthermore, the number of potential stimuli for studies is abundant, ranging from public criticism of a program, to spare research funds (which have to be spent before the end of the fiscal year), to a genuine need for the program administrator to know how the program worked out in practice. The following remarks will be confined to those initiatives which occur because the evaluation function is institutionalized somewhere in the politico-administrative system.

Legal obligation and parliamentary request Due to the outstanding role parliaments played in bringing about regular evaluations in most of the countries, it is not unusual for the imperative to carry out an evaluation to be laid down in the specific law or grant that authorizes the respective program. Examples of this kind of program-specific institutionalization and initiation include

- Economic Opportunity Act (United States, 1964)
- Compensatory Education Program (United States, 1965)
- Joint Federal-State Program of Regional Developmental Subsidies (FRG, 1969)
- Further Training Program in the Labor Market Law (FRG, 1969)
- Law of Continued Wage Payment in Case of Illness (FRG, 1969)

- Experimental Cash Public Assistance Program (DK, 1984)
- Law of Education and Training Allowance for Young Long-Term Unemployed Persons (DK, 1985)

In these mandated cases, the executive branch is obliged to produce a report to parliament after a certain period or at regular intervals; while the law is not always specific about the contents of the report, the task has been generally interpreted to mean evaluations. The practice of institutionalizing the evaluation in the authorizing legislation itself has been done in the United States by way of turning to sunset legislation as a general congressional policy. With this policy strategy, programs which have not proven effective in an evaluation expire automatically after a specific number of years. German Parliament has followed this device in a limited number of cases, although there is no juridical consensus yet whether a legislated program may be terminated at all.

Another mode in which parliaments become initiators of PE is the simple request for a periodic report from the executive branch. For instance, most of the two hundred regular executive-branch reports in the FRG (far from all of which are evaluations) to Parliament originate in standing parliamentary requests.

As evaluation studies almost always involve collecting and analyzing huge amounts of data, parliaments regularly ask the government bureaucracy to do the investigation, and often enough the actual research is further delegated to outside specialists (see the subsection "Internal versus External Evaluation Research" in this chapter). There is, however, an exception to the rule: where a parliament itself supports a large research staff and assistant offices, the initiatives will be more often directed towards these specialists in the legislative branch, as is the case in the United States. In these circumstances there are possible instances of parallel evaluations carried out in the executive branch as well as in these legislative support offices.

Initiatives of supreme audit institutions The national reports are not clear as to the extent the audit offices, which have been of paramount importance in institutionalizing evaluation, initiate and carry out evaluations themselves. Undoubtedly, the more the audit office has been the promoter of general evaluation activities in the executive branch, the more it will advise government and departments later on to evaluate specific measures, unless the department wants to evoke public criticism of the audit office. On the other hand, in the actual auditing process, evaluation studies are possibly carried out by the auditors themselves, depending on their expertise in this task. Both functions, advising the government to evaluate activities and carrying out studies of its own, are

fulfilled by the U.S. GAO with its methodology division. Across all divisions and responsibilities, the GAO conducts approximately 1,050 studies at any one time.

It is important to stress that only a small number of these studies should be considered evaluations. Financial audits, management studies, and investigations of fraud, for example, are all also included in this total number. Still, the number of evaluations would total several hundred. The Swedish NAB fulfills the dual function of advising how to do evaluations and of carrying them out.

Most other audit offices in Europe, with their traditions of several centuries, may be less prone to adopt the new terminology and therefore appear to have been rather inactive in this field until recently. In practice, however, their work might not be all that different (while the numbers are definitely smaller) from that of the U.S. GAO. Nevertheless, that entire programs are evaluated by European audit offices might be unusual, except for spending on defense equipment programs. However, the concept of policy evaluation seems to have spread to these auditing institutions recently.

Finance ministry initiatives Another source of initiatives can be the finance ministries or equivalent offices such as the OMB in Washington, D.C. Whereas the parliament and the general accounting office can both be regarded as institutions external to government (in the European sense), we now move to actors in the executive branch as potential initiators. Of course, the audit office reports are fed back into the budgetary process; that is, ex post analyses will have an impact on future budget appropriations and even budget proposals of the finance ministry. The relationship between evaluation and budgeting will change, if the finance ministry is in a position to request evaluations from the program administrators in the departments. As has been pointed out before, this is the case in the FRG. The closeness of this link between budgeting and evaluation depends on the frequency and regularity with which studies are requested; if it is a matter of bargaining between finance and program departments, it is likely that evaluations are sacrificed in order to minimize conflict. A structural solution to cope with this problem might be seen in the establishment of an Agency for Policy Analysis with the Dutch Finance Ministry, whereas the Danish system of annually rotating evaluations from policy area to policy area is a procedural device for linking evaluation more firmly to budgeting.

Obviously, the standing of finance departments varies from country to country. It remains to be seen if, in the context of the management initiatives, finance departments will assume as strong positions as the OMB in Washington has enjoyed. Again the question arises whether the

studies labeled "evaluations" which finance departments request for legitimizing annual budget proposals deserve this name despite their necessarily quick and short preparation.

Central political initiatives Taking the initiative to institutionalize evaluation procedures in government does not imply that the head of government, or the cabinet, or the cabinet office as the central political actors also initiate specific evaluation studies. Technically this is hardly possible for smaller programs or projects, unless these become a matter of political concern when public criticism arises or principles are at stake and are brought up by a minister. The suggestion to evaluate the respective program will, however, regularly be made by the respective cabinet minister and, therefore, will constitute a departmental initiative. Possibly, a cabinet office would suggest an evaluation owing to its intimate knowledge of departmental affairs.

What central political actors are likely to do is take the initiative to "assess" policies rather than programs; that is, they deal with broader policy or problem areas, which occasionally cut across departmental jurisdictions. They are interested in comprehensive analyses referring to these questions. The broader and the more unspecific the problem under consideration, the more likely is their treatment in commissions and committees. The inverse suggests specific program assessments could be done by in-house specialists. Examples of both, broad approach and central specialist evaluator, can be found in Britain: the former Central Policy Review Staff was an institution directly reporting to the cabinet, and it reviewed broad policy problems and developed new options. The Efficiency Unit followed this line after 1979; its effort to make the machinery of government more efficient involved, however, less substantive studies but rather organizational-reform studies aiming, among other things, at departmental self-evaluation.

Since 1981, the Dutch cabinet decides on the subjects (about one hundred topics through 1988) to be assessed by departmental study groups within the annual Reconsiderations Procedure. The focus of these studies is, however, not confined to questions of program efficiency and effectiveness, but involves organizational aspects, too. The OMB in Washington should also find mention here, as it carries out central budgeting functions and is a central government office. Obviously, the OMB is a central initiator, as it requests evaluations from the departments.

In conclusion, it might be maintained that at the center of government in most countries no initiatives are taken to evaluate specific programs and projects. The exceptions to the rule are the U.S. OMB, whose initiatives derive rather from its budgeting function than from its political staff function; the British Efficiency Unit, which initiates evaluations to the

extent they are part of the policy to improve management operations; and the Dutch cabinet, which takes specific evaluation decisions.

Generally one can state that the center of government, apart from emphasizing the need for evaluations and institutionalizing the function, engenders at best broad policy appraisals, which in and of themselves do not meet the formal requirements of a thorough evaluation. These assessments and appraisals, as they are often termed, are then prepared by central government staff units or by advisors and committees.

Departmental initiatives As was mentioned above, departments tend to resist self-evaluation; thus parliaments and their support agencies, as well as cabinets, for this reason have institutionalized or propagated departmental evaluations. Consequently, any of what may appear at a first glance as departmental initiatives are in fact externally induced. Examples are the Swedish central agencies which are responsible to government for evaluating their programs. Nevertheless, there are genuine departmental initiatives. The evaluation efforts in connection with the intervention programs of the 1960s seem to have originated in the respective departments; the more so the greater the affinity between program administrators and relevant outside researchers who dealt with the problem the new program was designed to cope with. In Washington, more than 80 percent of the departmental and agency evaluations in 1984 were internally mandated—in every second case by a top agency official, whereas only 9 percent were requested by Congress and 3 percent by the OMB. This distribution is typical for the rest of the nations and reflects the maturity of the U.S. evaluation tradition as well as the enormous personnel capacities of the evaluation units in Washington.

From a government-wide perspective these decentralized initiatives clustered in certain policy areas (education, employment, health care, developmental aid) and did not involve or cover the majority of departments.

In order to steady the autonomous intradepartmental initiatives or to initiate evaluations in those so-far inactive departments, departmental evaluation units were created. These units range from those which plan and design studies and write reports without necessarily doing the actual research involved, to those which take on all aspects of evaluations. Staff units (206) were observed across all U.S. nondefense departments. In some German ministries (Developmental Aid, Traffic, Press Office, Agriculture, Technology) evaluation units are established. In Denmark these units exist in the departments of Developmental Aid and Labor and the Directorate of Welfare. In Norway the Ministry of Development Cooperation created an evaluation division in 1983.

Making the deputy ministers in Canada responsible for departmental evaluations can be regarded as a functional alternative to the staff solution. In Sweden, expert groups attached to most of the ministries are involved in evaluation activities. In the Netherlands departmental studies are prepared under the Reconsideration Procedure with Finance Ministry officials participating in the work.

To summarize the answers to the question of who initiates evaluation studies in the policy process, one can probably say that in most political systems we meet a multiplicity of stake holders. This fragmentation largely fits the picture we had arrived at in the previous sections with respect to the process of institutionalization. Apart from external institutionalizations and recurrent initiatives, the executive branch itself takes initiatives to evaluate its own activities (to put it carefully). As to the central management units, initiatives can originate in the finance department and become part of routine budgeting. Central political actors, though, tend to restrict themselves to strengthening the inclinations of departments to undertake evaluations at all, while broad policy appraisals of the commission-type can hardly be regarded as evaluations proper. On the decentralized level of the departments, evaluation activities are increasingly preprogrammed by legislated evaluation requests. Furthermore, the evaluation function, partly in response, tends to be structurally differentiated from the function of the program manager.

Internal versus External Evaluation Research

Initiatives for evaluation are usually directed at actors other than the initiator himself, nor do those responsible for presenting evaluation results always carry out the necessary research for preparing the report. When we ask who actually evaluates a program, we find that evaluation research is only partly done within the structural units mentioned above. Rare cases of in-house research are complemented by an overwhelming amount of commissioned external research.

In-house research To carry out evaluation studies within the politico-administrative system requires at least specific jobs and at best professional policy analysts familiar with evaluation methodology. In the U.S. executive branch (1984), 1,179 professional staff members were occupied in evaluation units and consequently only one-quarter of the 1,689 studies in 1984 were conducted externally. In addition, at any point in time, the GAO is conducting another several hundred in-house PE studies.

The extent to which in-house research is carried out in government

departments might strongly depend on a government's organizational principles. In most countries the ruling principle is that departments should be as small as possible and serve the minister as his staff; under these circumstances the ministry would be badly advised to engage in research activities. What the ministry will do, however, is to formulate the policy conclusions drawn from a research report, whereas the research activity would either be delegated to governmental and quasi-autonomous governmental or nongovernmental research agencies, or commissioned to outside researchers. Federal government research agencies, for instance in the FRG, are numerous in technical policy areas (agriculture, traffic, commerce).

Of course, the picture is different where the task of the organization, say the GAO, is to evaluate departmental programs. Here the research activity is the core of the agency's mission.

External research As a rule the executive branch, thus, will transfer evaluations, which involve research activities, to specialists outside the administration. These can be either government-controlled institutions or independent centers, such as academic or commercial contractors (see Wagner and Wollmann 1986b).

Examples of government-controlled research agencies in the FRG include the Research Institute of the Labor Market Administration (founded in 1969), the Science Center of Berlin (1973), or most of the nationally important economic research institutes like the Deutsches Institut für Wirtschaftsforschung in Berlin. Sweden has a number of sectoral policy research institutes which, although being part of government, have a high degree of academic freedom. In this respect the statistical offices will also be potential research institutes to deal with evaluations. Government control or sponsorship does not imply that the research is scientifically dependent and possibly biased, but rather that government is in a position to ask the institute to investigate a certain question.

The availability of independent institutions reflects the applied character of a national social-science culture. While some countries give research predominantly to universities (particularly in education policy), others may be in a position to draw on independent institutions like the Brookings Institution in Washington, D.C., U.S.A., or on semiautonomous institutions like the Danish National Institute of Social Research (founded in 1958) or Denmark's Local Government's Research Institute on Public Finance and Administration (founded in 1975); furthermore, in the 1970s Canada founded the C. D. Howe Institute and the Institute for Research on Public Policy.

Where surveys (sampling, interviewing) are necessary to analyze a program, it is more likely that commercial research institutes become in-

volved. These institutes can be found in most of the countries under consideration, although their number naturally varies according to the size of the research market.

Administration of External Research

If evaluations are externally contracted, the interaction between researcher and administrator becomes problematic and is often dealt with in the literature analyzing the process of giving policy advice; I shall not embark on this problem but only deal with three aspects here.

Contracting evaluation researchers Has evaluation research become competitive? In principle, it is favorable for the mandating government official if he can choose between alternative external researchers; from a purely economical point of view, government will wish to contract the qualitatively best and least expensive evaluator. Political considerations may play a role, too, and induce the government to give a contract to an institute whose views are similar to those of the program administrator.

In practice, however, in most countries the supply side might be not so strongly developed that there are alternatives to choose from. It is indicative of the state of the art that the U.S. data show three-quarters of the four hundred research contracts (1984) following competition between contractors. In particular, in smaller countries one cannot expect an evaluation market big enough for equally specialized institutes to develop. In Sweden, therefore, most of the studies under responsibility of ministerial research delegations are commissioned to universities or public-sectoral research institutes.

If quality control of research is a problem in general, it is all the more the concern with sole-source contracts. One of the few quality-control mechanisms available for sole-source contracts is to make sure that the final results are published and thus open to public scrutiny and study by other researchers.

Financing evaluations The costs of evaluations are high and not seldom cost up to a million or more of the national currency; in particular in the United States evaluation has become a multimillion-dollar business (with 20 percent of the 1,775 evaluations in the United States in 1984 costing more than one hundred thousand dollars). Unfortunately, we do not have data from other countries. Often, the budget is not clear enough to allow inferences in this respect, or evaluation resources are hidden behind various items. At best the volume of government research could be determined irrespective of type.

The source of evaluation funds varies with the arrangements and structure of the evaluation system. With evaluation offices and in-house

research the costs are basically covered by the institution's budget. If research takes place external to the bureaucracy, the source of funds can be with the mandator of the evaluation, for example, parliament, or with the program administration. Furthermore, the funds can be dispersed over many budget items or centralized in a departmental, government-wide or parliamentary research budget. The more firmly evaluation is institutionalized, the more likely it is that there will be the allocation of specific funds, for example, to a departmental evaluation unit. An interesting variation of this arrangement has been devised in the United States, where evaluation funds are occasionally earmarked and stipulated within the legislation creating the program itself.

Dissemination of studies A central political aspect of the evaluation process is the extent to which the studies are circulated. Obviously, reports to parliament are important and can provide the stimulus for a broad political discussion about the future of a program. For this very reason, evaluations undertaken or contracted by government departments tend to be secluded from the public. The practice depends on legal norms regulating the freedom of information. Naturally, the degree of secrecy or publicity attached to this kind of information is determined by the national political and administrative cultures. It can be taken for granted that freedom in this respect is largest in the United States, Scandinavia, and the Netherlands, whereas Britain might keep reports, unless addressed to Parliament, even more secret than the FRG. In Germany researchers are considering whether to accept government contracts only if subsequent publication of research reports is safeguarded. Practically, however, the question of public access to research reports has never been a problem: it is impossible to keep secrets in Bonn.

Publication involves more than public access; reports would have to be printed. The costs of publication, though, has nowhere been a problem.

Concluding Remarks

In writing this summary I became increasingly aware of the fact that the availability of data on the national evaluation arrangements is itself an indicator of the state of the art in the various countries. The more firmly evaluation is institutionalized and the longer its tradition, the greater the propensity to analyze and even evaluate the evaluation business and to engage in meta-evaluation research. Therefore, the tentativeness of this summary reflects the still-shaky state of evaluation in many of the countries considered here.

This weakness pertains also to the very concept of evaluation. Where the evaluation function is firmly institutionalized and a tradition of eval-

uation research has developed, the term is employed in the narrow sense of methodologically controlled analysis of program strategies and impacts. In other countries, it has a rather broad connotation embracing various feedback mechanisms apart from research-based ones and a multiplicity of criteria beyond that of effectiveness.

I would not go so far as to maintain that the second wave of evaluations, which was primarily initiated by parliaments and the political executive putting the auditing institutions into the limelight, tended to water down the concept of evaluation. There are, however, strong indications that the further the evaluation function moves away from the departmental program manager, the more it becomes amalgamated with traditional auditing and review practices focusing also on input, output, and legality of implementation and not merely on strategy and impact of programs.

Note

1. The name of the BOB was changed to Office of Management and Budget (OMB) in 1970, as PPBS was intermittently called Planning Programming Evaluation System (Schick 1971).

References

- Adams, B., and B. Sherman. 1978. "Sunset implementation: A positive partnership to make government work." *Public Administration Review* 38, 78-81.
- Bulmer, M. 1988. "Social science expertise and executive-bureaucratic politics in Britain." *Governance* 1, 26-49.
- Nioche, J.-P., and R. Poinsard. 1985. "Public policy evaluation in France." *Journal of Policy Analysis and Management* 5, 58-72.
- Quermonne, J.-L., and L. Rouban. 1986. "French public administration and policy evaluation: The quest for accountability." *Public Administration Review* 46, 397-406.
- Rossi, P. H., and S. R. Wright. 1977. "Evaluation research: An assessment of theory, practice, and politics." *Evaluation Quarterly* 1, 5-52.
- Schick, A. 1971. "From analysis to evaluation." *The Annals of the American Academy of Political and Social Science* pp. 57-71.
- _____. 1988. "Micro-budgetary adaptations to fiscal stress in industrialized democracies." *Public Administration Review* 48, 523-33.
- Wagner, P., and H. Wollmann. 1986a. "Fluctuations in the development of evaluation research: Do "regime shifts" matter?" *International Social Science Journal* 205-18.
- _____. 1986b. "Social scientists in policy research and consulting: Some cross-national comparisons." *International Social Science Journal* 601-17.
- Ysander, B.-C. 1983. "Public policy evaluation in Sweden." The Industrial Institute for Economic and Social Research. Working paper no. 106.