

UC Berkeley
IURD Working Paper Series

Title

POLICY EVALUATION: THE UNCERTAIN GUIDE

Permalink

<https://escholarship.org/uc/item/7vd3c9fd>

Journal

Institute of Urban and Regional Development

Author

Teitz, Michael B.

Publication Date

1978-09-01

POLICY EVALUATION: THE UNCERTAIN GUIDE

Michael B. Teitz^{*}

Working Paper No. 298

September 1978

^{*} Professor of City and Regional Planning
University of California, Berkeley

This paper was prepared for delivery at the 15th Biennial Congress of the Royal Australian Planning Institute, Sydney, August 28th - September 1st, 1978. My thanks for suggestions and comments to Melvin M. Webber and Lyna Wiggins.

INTRODUCTION

During the past ten years, we have seen in the United States an extraordinary increase in resources and effort devoted to the systematic evaluation of public policies and programs. Clearly, evaluation of policies in itself is not new. Effective rulers have always used information and experience to design policies and modify them in action. What have changed are the scale of evaluation and the methods whereby it is done. Unprecedented amounts of money and expertise are being allocated to the determination of whether policies and programs do what they are intended to do, whether they are too costly, and whether they have other consequences detrimental to public concerns.

Carefully evaluating the desirability of public actions is, on the face of it, totally unexceptionable. Who can be against the idea that decision makers should know more about the consequences of their actions? Yet, the issue is not that simple. We are coming to understand how ideology and interest shape our perceptions and subtly influence what we take to be objective reality. Who evaluates, what is to be evaluated, and how questions are framed may set an agenda for public discussion that excludes important concerns. Our technical capacity for analysis may look impressive yet be insufficient to deal with the complexity of the questions that we seek to answer. The result may be simplification or distortion of reality. The process of evaluation itself may seriously interfere with or change the way in which policy is made, not always for the better. In short, evaluation is neither neutral nor invariably beneficial.

The very fact that the RAPI Congress has a day devoted to policy evaluation is some indication of the power of the social and institutional forces in our societies that underlie an evaluative stance toward public action. My sense is that we will be living with evaluation into the indefinite future. In order to gain as much as we can from the institutionalization of this activity, we should be both knowledgeable and

critical about it. This paper is intended to provide a general framework for discussion of policy evaluation in specific areas such as transportation or taxation. For this purpose, I will discuss five principal topics:

- Defining policy and program evaluation.
- Evaluation as a societal response.
- Methods for effective policy and program evaluation.
- Benefits and costs of policy evaluation.
- The relationship between planning and policy evaluation.

This is a large agenda. However, a broad perspective seems appropriate at this stage in the development of this field.

DEFINING POLICY EVALUATION

In current usage, a number of expressions describe evaluative activities in relation to public actions. They include policy evaluation, program evaluation, policy analysis, impact analysis, evaluation research, and program planning. I do not want to attempt a definitive classification of these terms. Their usage is too varied for that. Rather, I want to start out with some important themes and distinctions that run through the topic as a whole.¹

In the United States, we have seen the rise of what might be termed an evaluational ethos in public policy. It takes many forms, but basic to all is concern for the impact and effectiveness of policies, programs, and actions. That concern is not simply for the direct or mandated objectives of policy, but also reflects the realization that unanticipated side effects may cause great problems and cost. With this realization comes rejection of the idea that, because a policy or program has worthy objectives in itself, it is necessarily in the public interest. Both public and private sector actions may have effects broader than their

¹There is now a very large literature on program and policy evaluation. Reviews may be found in Glass (1976), Guttentag (1977), Dolbeare (1975), and Teitz (1978). An excellent introduction to program evaluation is Weiss (1972).

immediate goals, so I think that we are being forced toward new and more comprehensive forms of accountability.

Forms of Accountability

We seek increased accountability for decisions that have problematical impacts. But, accountability is not, in itself, a simple objective. At least four types of accountability can be identified in the movements that are now occurring. The first is accountability of policies and programs for the achievement of their stated objectives. Legislators and their constituents are asking hard questions about why schools apparently do not teach children to read, prisons do not reform, police forces do not inhibit crime, and social services do not alleviate distress. We are asking whether the basic theories upon which our policies and programs have been designed are actually correct, or, more succinctly, are programs effective? To answer this type of question is very difficult indeed and calls for basic evaluation research on the relationship between what programs do and their ultimate impact. For example, we have spent about \$1.6 billion to build the BART system in the Bay Area. Now, we are spending another \$7 million to find out that it will do virtually none of the things it was supposed to do, and is paid for by the poor and ridden by the well-off [Webber (1976)].

A second form of accountability deals with the question of efficient resource use. In California, we are now undergoing a tax revolt against the cost of government -- the infamous Proposition 13. However, it should not be taken as a pure ideological rejection of public intervention. Polls show that people are not against public services as such, especially those from which they benefit directly. Rather, the political pressure springs from a sense that public services cost more than they should, and that inefficiencies are perpetuated in the interests of those who manage the public sector. Whether true or not in any ultimate sense, this perception leads to calls for analysis and evaluation of the cost and efficiency of public services; which is not the same as asking whether they are also effective.

A third major accountability issue is the external or side effects of public and private decisions. Political sensitivity to such effects has resulted in a proliferation of evaluation requirements written into laws governing both public and private actions. At this time, any major piece of legislation being considered for support by the administration in Washington is subject to reviews for environmental impacts, impact on the family, impact on the budget, and urban impact, particularly with respect to older cities' and metropolitan areas' economic structure and employment. At the project level, the environmental impact statement is the predominant form of assessment, but not the only one. To site a new coal-fired power plant in the U.S., depending upon the state, may require analysis of its potential impacts on the environment, including air and water quality, endangered species, land use, and socioeconomic structure that involve as many as fourteen separate agencies. The lead time for such a plant is now 8 to 10 years.²

A fourth concern is accountability to specific population groups affected by policies or actions. This concern might well be considered part of the externalities issue, but I separate it because in the U.S., perhaps more than elsewhere, evaluation of the impacts of policies on disadvantaged groups has become institutionalized. For example, access to jobs for minorities and women is a standard dimension for the evaluation of any major program, for example, the BART evaluation. In another sense, the idea of accountability is reflected in requirements for participation by population groups in program decisions likely to affect their welfare. Such participation is not always effective; often it is consciously manipulated. Nevertheless, it is an integral part of the evaluation structure. Requirements for impact assessments have also indirectly promoted a form of participation. Environmental impact statements themselves are often of dubious quality or ignored by decision makers. But the fact that they are required by law has given interested groups the legal right to intervene in decisions through challenges to EIS adequacy. The result has been to delay initial approvals of projects or to stall them after formal approval, a significant veto power.

²For a discussion of this problem, see Nagel (1978).

These diverse forms of accountability suggest that there might be varied organizational forms for evaluation. I will not attempt to describe the structure of agencies that are responsible for these various forms of accountability. To tell the truth, I do not think that anyone can comprehend the full range of federal, state, regional, and local agencies that carry out evaluation, either as a principal function or in conjunction with responsibilities for planning or regulation. Instead, I would like to focus on two broad distinctions in evaluation: between retrospective and anticipatory, and between expert and adversarial.

Forms of Evaluation

We may distinguish broadly between retrospective evaluation in order to draw conclusions from the operation of programs and policies that have been implemented, and anticipatory evaluation of designs and alternatives for policies, programs and projects that are being proposed.

Retrospective evaluation looks at policies and programs after the fact. Its objective is to provide feedback information on how programs have performed, presumably so as to improve future decisions about them. Such evaluation has been legislatively mandated in various ways, for example, through the creation of formal planning and evaluation units within Federal agencies, or by Congress setting aside some percentage of funds in a program's budget to be used only for evaluation. It is carried out directly by subagencies in the Federal government, for example, the Office of the Assistant Secretary for Policy Development and Research in the Department of Housing and Urban Development; by independent audit or evaluation agencies, such as the General Accounting Office, or the Congressional Budget Office, which reports to the legislature; or it is done under grants and contracts by a huge variety of firms and nonprofit institutions acting for government agencies, or sometimes independently.

The most common model for all their work is applied science -- to find out as objectively as possible what has happened, compare it to what was intended, and to draw conclusions about it for the future. But the objectives within that mandate vary greatly. Writers on evaluation have

especially distinguished between outcome and process oriented evaluations (otherwise called summative and formative). Where the evaluators focus on the question of outcome effectiveness -- did the program or policy achieve what it set out to do -- they tend to behave most like scientists, concentrating on the problems of research design, accuracy of observation and measurement, and validity of inference from their data. The medical model for testing the effectiveness of treatments, with its apparatus of statistics and double-blind experimental designs comes closest to this type of evaluation in its pure form. There tends to be little direct emphasis on policy or program redesign; the logical inference from a clearcut result is that a program should continue or be terminated. Unfortunately, this rarely happens, either because the results are not clearcut or because the program has other political reasons to exist than effectiveness alone. Furthermore, the emphasis on reaching an analytical conclusion seems somehow antithetical to treatment of the problem as one of policy or program redesign. As a result, this type of evaluation, although often the most elegant and intellectually interesting, does not tend to have clearcut policy impacts. It does on occasion stimulate enormous debate. Evaluations of the effectiveness of school-bussing for reducing the effects of de facto racial segregation, for example, have produced much argument but have not settled the matter at all.

Another mode of retrospective evaluation emphasizes analysis of processes within a program's operation. This style may involve more informal means of analysis, perhaps relying on skillful observers, and also focuses on the processes at work within the system. It therefore tends to be conducted less dispassionately. On the other hand, identification of problems and modification of the program are central concerns, so the evaluation process comes to be less cataclysmic in its potential impact and less threatening. By the same token, it may be more easily coopted by the program's professionals, clients, or advocates, with whom the evaluators are in frequent and close association. Where outcome oriented evaluation tends to deal with the question "to be, or not to be," process oriented evaluation asks "What's going on here?"

I might illustrate these differences by reference to a recent study for the U.S. Department of Health, Education and Welfare to evaluate 11 demonstration projects for dealing with child abuse. In the course of the evaluation, we attempted to find out how well the programs worked (in relation, say, to reincidence); how much different treatment modes cost in relation to effectiveness; and who was served. At the same time, full understanding required the evaluators to look at how the programs worked, for example, to consider worker "burn out," the phenomenon of staff simply losing their desire or care about work.³

It is easy to become immersed in these somewhat arcane differences, but let us return to the distinction between retrospective evaluation of existing programs and anticipatory evaluation. Whether outcome or process oriented, the types of evaluation that I have been discussing remain firmly grounded in the idea of accountability after the fact. Treating government as a continuing activity, we apply whatever means that seem appropriate to find out about its performance in action. Contrast this with an evaluative stance that seeks to find out what the result will be before the fact -- the classic design or planning analysis question. Anticipatory evaluation is now as widespread as evaluation after the fact. However, its organization and institutionalization differ considerably.

Anticipatory evaluation deals with a future decision. Some choice will be made and we would like to know its consequences. As I have suggested earlier, this is likely to require analysis of impacts, and the evaluational problem is to determine which impacts, who does the analysis, and who is represented in the decision process. Such a situation is likely to be inherently one of conflict among interests, a familiar situation to planners. It seems to me that in the U.S. we have been evolving over the past years a new structure for dealing with evaluation, especially at the project level. It has been a painful and often incoherent process, the end of which is not in sight. The most notable feature of this process appears to be the establishment of legislation and of formal ways in which proposals for policies and for public and private projects may be

³Like much of the substantive research in program evaluation in the U.S., this work is mainly in the form of reports to a governmental agency. See Berkeley Planning Associates (1978).

scrutinized and affected by groups or interests that may suffer from external effects.

Two elements are important for this process. The first is a structure of requirements for estimation of impacts of proposals on a variety of dimensions. I have previously mentioned the requirement for environmental impact statements for any project of significant scale. Another illustration is the requirement that new buildings meet standards of energy efficiency in performance. To a substantial extent, the responsibility for the analysis of performance or impact has been placed upon the agency or other entity, for example, the developer, responsible for the proposal. Review and approval of the adequacy of impact analyses both as to quality and as to the seriousness of the impacts revealed falls on other agencies, for the most part, public agencies with planning and review powers for specific functions or areas. In California, the number of such agencies is large and their responsibilities intersect in complex ways. They include, for example, the Coastal Commission for proposals that affect the coastline, the State Energy Commission for energy issues, the State Public Utilities Commission for cases that fall within its regulatory jurisdiction, metropolitan areawide planning agencies for capital improvements, and so forth. The siting of a major facility such as a liquefied natural gas terminal must pass through a long, drawn out process of analysis and review by such agencies before it can be approved.

The idea that the impact of proposals should be examined and evaluated by an external agency is not necessarily very powerful. What makes it socially and politically potent is the introduction of legitimate paths for groups that have some interest in the decision to affect the evaluation process. Historically, the principal paths were through the political system, either by access to legislators or the executive. For politically weak groups, these means were quite limited. In the U.S., they have been augmented by the large scale use of litigation and court ordered intervention. Challenges to the adequacy of environmental impact statements or to the legality of subsequent actions in the light of EIS conclusions have been most prominent ways in which public decisions have been forced

to take account of considerations other than the direct objectives of large projects. Such challenges have become possible in good part because of the recognition of external impacts on air quality, endangered species, and relocation of individuals and families. Permission was denied to fill the huge \$100 million Tellico dam in the Tennessee Valley because of the threat to the existence of the previously unknown habitat of the two-inch-long snail darter fish.⁴ Another dam project in the state of Maine was blocked upon discovery of the habitat of an obscure plant known as the furbish lousewort. These are only the most publicized cases in a much longer list. Proposals to develop major industrial plants in California have been blocked on the basis of air pollution and negative impact on public health. Increasingly, we are seeing challenges to further industrial development, even with substantial unemployment, as analyses of the relationships between emissions and disease become available. In such cases, we often find a combination of private and public actors with government agencies ranged on both sides of the issue. And initiation of suit may not only rest with private groups. Thus, the federal Environmental Protection Administration may find itself suing the State of Texas in order to force compliance with federal air quality statutes in relation to power plant siting. Party to such a suit might be state agencies, private environmental groups, and power companies.

Adversary proceedings in evaluation are not restricted to situations in which policy or program decisions have yet to be made. In housing, the federal Housing and Urban Development Department has been successfully sued on the basis of its non-performance in programs that are statutorily mandated. Similarly, state mental retardation programs in New York have been sued to force upgraded standards of care and treatment.⁵ Nevertheless, requirements for impact assessments do make the situation in the anticipatory

⁴The full saga of the effect of species protection in the U.S. has yet to be written. See Carter (1978).

⁵Glazer (1978) discusses the problems in this form of intervention. An opposing view appears in Gaylin, et al. (1978), especially in Gaylin's own essay.

case rather different from that of most ongoing programs, where the evaluation process has been internalized and it is not easy for other participants to gain access.

To summarize, I have suggested that evaluation has become a very important function in the U.S., both as a means for assessing programs' and policies' performance and cost, and as a way to ensure accountability for the external effects of proposed programs. Evaluation in the United States seems to be conducted according to two models. On the one hand, we have what might be called the expert model. Here, evaluation is seen as an analytic process in which scientific methods are employed to determine as precisely and comprehensively as possible the effectiveness and cost of policies and programs. This work will normally be done by an expert group in an agency for that purpose. It is both retrospective and anticipatory. It tends to be closed with respect to politics; there is one client, preferably the decision maker, and the evaluation is a way to reduce conflict and argument by determining the "facts" which will be accepted by all parties. On the other hand, the adversary model explicitly opens the evaluation process to argument. It is based upon a legal right rather than upon a notion of beneficent government. Experts may be employed and analysis may be very sophisticated, but the results are intended to be argued in either a legislative or judicial process. This model tends to appear more often in anticipatory situations. Its politics are more open because participation is legitimate for groups outside the structure of government itself.

I will explore the implications of these forms further, but first we should pause and ask why the idea of evaluation has taken hold so strongly.

EVALUATION AS A SOCIETAL RESPONSE

Thus far, I have suggested that the overt purpose of evaluation is to increase accountability with respect to objectives and impacts, and that its nature and organization vary depending upon the context in which it is undertaken and the political process of which it is a part. There is another way in which to ask the question "why evaluation?" that I think is also worth exploring. Our understanding of how best to proceed with any

important social function may be improved if we look beyond its immediate purposes to ask why that function seems to be demanded at this particular time. After all, major decisions have been made as long as there have been societies, but governments have not always sought to determine the impact of policies by analysis.

It seems to me that two interrelated social characteristics have a lot to do with the rise of an evaluational ethos. The first is the growing interconnectedness of modern societies, a trend that has been accelerating rapidly during this century. Related to this fundamental trend is a second characteristic, namely the growth of government and of collective decisions outside the market or traditional social structures.

Ages and societies have their own distinguishing forms of thought, behavior, social organization, and production. If there were any unique intellectual mark of our times, I think that it is the idea of interrelation or system. The nineteenth century can be seen as reductionist in its temper, seeking to break down things into their understandable elements; the twentieth century seems more centrally concerned with structure and interrelations. Our technology's greatest achievements are systems of increasing scale and complexity, from power distribution to satellite communications. Perhaps the most gripping scientific discovery of our times determined the structure of DNA, not its composition which was already well established. In biology, the study of the organism within its structure of ecological relations has reemerged as a central problem, and the field of ecology has taken on new life. There are many social and economic analogs.

Prevailing ideas reflect the society in which they occur. Modern industrial society is interconnected more profoundly and at a far larger geographic scale than any that preceded it. Just as worldwide communications systems have reduced the time-lag in transmission of information from years to virtually zero for critical events, so the integration of production and consumption has become worldwide. I need hardly remind Australians of the degree to which their economic relationships with the world have changed over the past twenty years. Within societies, too, old patterns have changed. The past three hundred years have seen the steady

removal of functions from the family unit; first production of goods, then services such as education and health care, then social support services in adversity and old age, and finally, perhaps a good part of child care itself as women increasingly enter the labor market. In each instance, the individual or household becomes more dependent upon networks outside its immediate control and more subject to the effects of decisions at a distance. As we depend more upon extended systems for livelihood and support, so we become more vulnerable to changes in other systems that affect them.

The costs of the interconnected society are high, especially the psychological costs of the loss of community and of alienation from work and society. As a result, the flow of change may not necessarily be all in one direction. The suburb as it has evolved in the U.S. has a strong element of search for a smaller, simpler community, albeit one in which work is controlled by a multi-national corporation and entertainment is packaged a continent away. Some people move further away; in the U.S. there appears to be a reversal of the trend of three hundred years as small cities are growing faster than metropolitan areas. Yet their ability to do so depends in good part on the means for national integration afforded by our communication and production systems. There have been arguments for the virtue and necessity of smaller, less vulnerable self-contained systems, whether in the form of utopian communities, solar power systems, or organic toilets. I do not want to argue the merits of these viewpoints, only to emphasize the complementarity of such movements to the central theme, namely, that we do live in societies that are internally and externally ever more interdependent, and that in such systems, changes in one sector will give rise to impacts in others at an increasing rate.

What are the implications of life in the systemic mode? In economist's terms, we are increasingly subject to externalities. The full social costs of some action may be neither perceived nor borne by those deciding to undertake that action. As they become more aware of this state of affairs, people will either attempt to insulate themselves from the effect of others' acts, or try to control them in such a way as to minimize the negative consequences for themselves. Since so many decisions are essentially irreversible -- the destruction of a species being only the most

obvious -- there is a strong incentive for creation of means for the prior assessment of and accountability for impacts beyond the direct objectives served by a policy or program before it is actually carried out.

Although there may be voluntary means to achieve such assessment, for the most part this kind of activity must be mediated by government. Growing societal interconnectedness, then, tends to stimulate the growth of the public sector in at least two ways. On the one hand, the older, self-contained institutions for maintaining the social fabric are no longer capable of performing their functions and are replaced or supplemented by new forms of service system. Insofar as these are collective, we may expect that they will be supplied through some version of taxing and public provision. On the other hand, interdependency of decisions requires ways to assess indirect as well as direct impacts, together with a means for resolving conflicts. In these two strands, I think that we can see the sources for evaluation as it has recently developed.

The growth of the public sector in industrialized societies is a familiar theme that I will not elaborate upon. Despite variations from country to country in the way in which particular functions are handled, and changes over time in the rate at which the process has occurred, there can be no doubt of the extraordinary increase in the amount of resources whose use is determined through public decisions rather than through individual choice. In the U.S., in 1951 about 76% of GNP was spent by the private sector, and 24% by government, of which 17% was federal and 7% state and local. By 1975, in a vastly greater output, the private share was 64%, and government disposed of 36%. The federal share had grown from 17% to 21%, while the state and local share more than doubled from 7% to 16%. In my opinion, these figures underestimate the total of what should properly be called public sector expenditure. However, the raw figures are sufficient -- both in proportional and absolute terms, the scale of expenditures determined through the political system has grown dramatically.

Some of the consequences of this situation have become painfully evident. U.S. governments at the local level are experiencing a severe fiscal crisis as their sources of revenue are unable to keep up with expenditures.

New York City has been close to bankruptcy. Reliance on the property tax has precipitated in California, and to a lesser extent elsewhere, a strong taxpayer reaction. But deeper, less obvious, changes were occurring well before the spectacular fiscal fireworks. Determination of how to spend public funds has grown increasingly difficult: people simply do not know what works or how new functions should be undertaken and the scale on which expenditures are made simply staggers the imagination. Both executive and legislative branches of government have adjusted in several ways. Evaluation and management have become central concerns. New organizational units have been formed that are intended to provide better information on program performance and efficiency. Many programs include mandated portions of their budgets set aside for evaluation. Others are subject to review by general evaluation agencies. An evaluation bureaucracy has been created to administer this function, and with it has grown a field with research, journals, training programs and all the familiar institutions that seem to follow.

The model of evaluation that best fits this situation is that which I described earlier as expert and closed with respect to politics. The institutions, methods and style of work that have evolved reflect this. Evaluation is a government activity that is carried on within the structure of administration and decision making. Information about the cost, effectiveness, and impacts of programs is determined by means of an analytical apparatus or by expert judgment. It is, in short, what we might normally think of when we think about evaluating a program or policy.

But, at the same time, there remains the other strand that I have described -- evaluation as an adversary process in which determination of impacts is in good part left up to the advocacy of the interests that are affected. They may employ analysis or expert judgment, but that is ultimately subject to contention as a part of the evaluation process itself. And the process is necessary because the conflicts are not resolvable by consensus.

Let me summarize this argument. The growing complexity and interdependence of modern industrial societies, with its concomitant interaction among decisions, implies more public sector activity to maintain

the society and to resolve conflicts. The growing cost of government, together with uncertainty about what will work, provide the impetus for demands for better information about programs' impacts, effectiveness, and efficiency, in short, for rationalization. The typical result is a mode of evaluation that is rationalistic in nature, depending upon expert knowledge and attempting to incorporate within itself the social conflicts of interest that are inherent in any policy. At the same time, conflicts of impacts promote a style of evaluation in which analysis is present but subject to challenge and refutation by groups that feel their interests at stake and intervene by right. Each of these is a response to the central dilemma of evaluation of public action, namely that it must combine analytic determination of impacts and at the same time face the political implications of its results. I would like now to explore the meaning of this dilemma for the methodology in evaluation.

METHODOLOGICAL ISSUES IN EVALUATION

One of the foremost figures in evaluation research in the U.S., James S. Coleman, has said that ". . . there is no body of methods, no comprehensive methodology, for the study of the impact of public policy" (1975, 19). That is a strong statement, but probably accurate. What we have is the eclectic adaptation of methods drawn from disciplinary research in social and physical sciences applied to problems of policy. This situation Coleman suggests may be due in part to the fact that governments have only recently begun to ask questions about the impact of policy, and in part to the fact that work in fields commonly called policy sciences has been predominantly directed toward the construction and testing of theory and the elaboration of knowledge within disciplines rather than toward information directed toward policy decisions.

A central methodological problem in evaluation is how to identify and measure the effects of policy decisions, or to predict them in the future. Up to now, we have relied primarily upon judgment or statistical methods, the latter usually some variant of experimental design. The classical assumptions of design of experiments can rarely be met in the real world in which very little stays constant. So evaluators have

struggled to isolate impacts as best they could by careful analysis of time series and cross section data, using multivariate analysis to attempt to account for the influence of extraneous variables that could not be held constant. But the task is very difficult. A feature that has been noticed by a number of people who study evaluation has been the large number of instances in which no impact whatsoever could be detected. Does that mean that none occurred? Or that we were looking for the wrong impacts? Or that we had the wrong indicators? Or that the impacts were perhaps there but drowned out in the noise produced by everything else that was going on at the time? It is frustrating not to be able to answer these questions.

One logical way to circumvent the issue would be to design controlled social experiments to reveal whether policies might work. Donald Campbell (1971) has advanced the concept of an "experimenting society" which incorporates an essentially scientific ethic of design, trial, and evaluation in an evolutionary process. The idea of social experiments is attractive in some respects. In an appropriate institutional structure, such a process would constitute a societal feedback mechanism. Indeed, in the U.S. during recent years there have been a number of experiments that explicitly attempt to meet rigorous design criteria. Among them are a negative income tax, housing allowances, and school and childcare vouchers. But the idea of the experiment is deceptively simple. It assumes a political context willing to take the results of the experiment as accurate and impartial; an organizational setting in which existing interests will not attempt to subvert the experiment for their own objectives; a consensus as to the variables that should be tested by the experiment and the allowable forms of intervention; and a willingness to pay the high costs involved. In short, it looks like the expert model of the detached evaluator raised to a new level of power. Experience so far with experiments suggests that they are indeed valuable. From the housing allowance experiments, we are learning a lot about housing markets. But even when funded in excess of \$50 million, the experiment is insufficient to tell us what would be the impact of a full scale housing allowance program. Much the same could be said about the negative income tax experiments. Economists have been

unable to agree on the effect of the subsidy on people's propensity to stay in or drop out of the labor market. Any ultimate decision is not likely to be made on the basis of information developed during these experiments, although they will obviously shape the way in which some people see the problems. The use of experiments cannot solve the political problem of evaluation by removing questions of fact from those of value.

Some evaluators have been developing a methodological approach that is grounded in a very different style from that of the detached observer. If improvement in the capacity of service delivery agencies to perform their tasks is the objective, then a more appropriate model might be that of the clinician who works closely with people in an agency, helping them to perceive themselves better and to change in ways that will be productive. A tradition of work of this kind in management counseling in the private sector has not been applied very much or very effectively in government. Clearly, such a style is more suited to incremental change within an institution that has an established role and set of relationships with the world at large. The political problem in this instance is more internal to the agency. Whoever is the client, the evaluator ultimately is likely to be seeking ways for the resolution of interpersonal or intergroup conflicts within the agency so as to enhance performance. Appropriate methods and skills tend toward the clinical, that is, careful observation, participation in a limited degree, structuring of situations and behavior so as to generate productive tensions and resolution of differences. Analytic methods and data may be used, but they are subordinate to the main task of intervention.

Let us step back and consider these methodological styles for a moment. Each style represents an accommodation to basic problems of evaluation. I have suggested that the meta-issue for evaluation is how to resolve the tension between research and decision, the necessity for political integration of the evaluation function. The expert model resolves this tension by a combination of comprehensiveness and positioning. We try to take into our analysis all the relevant variables that represent concerns of significant affected interests. In order to ensure political impact, the evaluator tries to position himself as closely as possible to

the key decision maker. The clinical model tends to show up in more limited situations, such as within organizations. It resolves the tension by being more developmental than analytic, emphasizing a role that is mediating in the sense of using observational technique to help participants see better what is at issue, and human relations techniques to help them resolve differences without falling into positions of mutual loss. The resolution of the tension, then, is achieved by redefining the evaluational function from the global assessment of benefits and costs, to enhancing the capacity of participants in a situation to do that for themselves.

Neither of these models quite fits the situation in which evaluation is an adversary process. Technical analysis is necessary, but it is far more likely to be employed in the service of an explicitly recognized point of view than to claim to be comprehensive. Adversarial differences need to be resolved, but that resolution is not likely to be controlled by an evaluator since the political and economic interests involved are likely to be substantial and recognized. So far, in the U.S., the mediation function seems to be handled by the courts and by legislatures and special commissions, all within the conventional political decision system.

Ultimately, I do not see that there is any simple resolution to the problem of evaluation structure and method. Despite objections to one or other forms of evaluation, we will probably continue to see the development of methodology in all the forms that I have described. Evaluation is a socially necessary function in modern societies, but its form and methodology will vary with the political system and the kinds of issues being addressed. Perhaps the most important lesson to be drawn is that evaluation, like other functions, should be carried out in a way that fits its environment. Trying to force a single model of policy evaluation onto a context to which it is not suited can only exacerbate problems that evaluation itself creates. And those problems may be significant.

BENEFITS AND COSTS OF EVALUATION

Ironically, but not surprisingly, there has been little assessment of evaluation itself. Debates about the conclusions or methodological quality of individual evaluations have occurred. Yet there has been remarkably little assessment of the effectiveness and cost of evaluation as a public function. Almost without our realizing the fact, the idea of evaluation has become embedded in public rhetoric and, more importantly, in budgets and organization charts. Its advocates have been busy building their careers and arguing about the merits of various methods. Its opponents have largely been confined to those who would cut public budgets on any ground, and those who are ideologically opposed to the expert model.

Yet, there are good grounds for viewing evaluation as a mixed blessing. Almost every positive attribute has its negative side. Consider, for example, the benefits of better information that is based on a full exploration of the costs and outcomes of a policy. In a decision process that is capable of accepting and using such information, we should expect better choices. But information is not costless. For complex programs and policies, the cost of finding out how they work and determining their immediate effects, not to say their long-run impacts, may be prohibitive. Lawrence Lynn (1977) has estimated that in the U.S., evaluatory information now costs the federal government in excess of \$1 billion (that is, \$1,000 million) per year. Does the public get its money's worth for all this effort? I am not sure that anyone can say, although there are many people who would assert not. One U.S. senator gives a "Golden Fleece" award each month to the most absurd expenditure that he can find. Evaluations are not absent from his list. More seriously, I know from experience that evaluation research is often duplicated or ignored. Yet that must be true of any information gathering process; search implies that we do not know exactly how to get to where we want to go. Proponents of evaluation point to the enormous savings that can be realized if ineffective programs are eliminated or effective ones promoted. Yet, so far, there is not much concrete evidence of such savings. In this matter, we are neatly caught on one of evaluation's own most difficult problems because we have no comparable situation without evaluation against which to judge.

More disturbing than the issue of cost are the criticisms of evaluation that stem from its attempt to be comprehensive and open. It has been suggested by Pressman and Wildavsky (1973) that the more steps and participants in any decision process, the smaller the probability that anything will be implemented at the end. Insofar as evaluation tends to increase the numbers of legitimate participants, we should find delays, and indeed that is the case. Adversary processes are particularly subject to delays and vetoes in a pluralistic and federal system. As the number of impact statements and reviews increases, so does the lament of those who would initiate projects and programs. Again, it is difficult to put any exact measures on these effects. At the project end of the spectrum, builders have charged that housing costs in California have been increased by as much as 30% because of the number of required permissions. Major projects, in particular, can be blocked for long periods far more effectively than formerly. Last month, construction began on a redevelopment project in downtown San Francisco twelve years after site clearance as it emerged from a cobweb of court suits over environmental and social impacts. The counter argument, of course, is that major projects or policies, such as the decision to build or not to build the supersonic transport, may be irreversible, extremely costly, and should not be undertaken without the fullest exploration of consequences. Attitudes to the costs of evaluation may well be more a reflection of peoples' ideology and perception of technological change than they are the result of the analysis of the costs and benefits. Personally, I feel that adoption of policies without giving full thought to their consequences is likely to be both costly and inequitable in a modern, interdependent society. So, I see delays as necessary. But I also believe that we need to develop institutions that can handle the conflicts better than we do at the moment.

Indirect effects of evaluation may prove to be even more important over the long run. For example, consider the question of the effect of a strong evaluation process on new programs. Although we may be able to avoid waste by early review, premature evaluation and discontinuation may be more serious in the longer term. One reason for this is the effect of a formal evaluational process itself. As it becomes institutionalized, evaluation may also become a procrustean bed that forces programs into

a particular mold and inhibits innovation. For social policies, changes are often regarded as questionable or even absurd when they first begin to appear. Evaluation as an institutionalized function is not likely to be capable of dealing with such change. This argues, I think against a single model of evaluation, or else as Coleman has suggested, for multiple evaluations of policies to be done simultaneously. The latter is quite unlikely on the grounds of cost, since it will be seen as duplication. In any event, it is no guarantee that other points of view will be heard.

What I have been saying here challenges but, I think, does not contradict my previous description of evaluation as an activity that is very much the product of modern society. Like any other emergent social function, evaluation precipitates its own problems and has its own contradictions. We are trying to find ways to deal with societal problems that are much more interdependent and at a larger scale than anything previously experienced. No one said that the task should be easy.

PLANNING AND POLICY EVALUATION

Much of what I have said must have a familiar ring to planners despite the differences in style and substance. In many respects, policy evaluation faces the old planning dilemmas. Planners have struggled with the question of whether it is possible to construct a field that has professional expertise and advances a broad public interest conception of welfare in a world that is essentially political with real conflicts of class or group interest and individual welfare. Planners can perhaps look at the process of development of policy evaluation and see their own predicament more clearly.

In some basic respects, planning and evaluation in the anticipatory mode are also quite antithetical. Let me illustrate this. One way in which we might see development planning is as a means for reducing uncertainty about conflicts and external effects in a very important sphere of decision, namely physical development. The coordination of development in time and space also has other objectives, for example, operating efficiency, but avoiding conflicts of use has been a central part

of the rationale for urban planning in its modern form. If every substantial project must be evaluated, as it were, from scratch, the planning process itself is profoundly altered. In the U.S., there are good reasons why this might be necessary. Plans have not been responsive to issues that have become more important in recent years. People have become much more sensitive to the impacts that public actions may have upon them. Planners have been seen as the tools of development interests, whether aware of that fact or not, and the results of planning seen as quite negative. Nevertheless, the idea of planning as a means of avoiding impact conflicts, rather than a process for reinterpreting and resolving them in every single instance where they occur, has great power. At this point, we seem to have recognized many of the deficiencies of the conventional planning process, but to meet them we have constructed a structure that is responsive but functioning only case by case and often excruciatingly slowly.

How we will resolve this problem remains to be seen. To some degree, planners have been moving into the new structure of regulation and assessment as professionals. This is especially true for major siting functions, for example energy facilities, and for specific areas of concern such as coastal zone development control where the adversarial model has established itself strongly. In these contexts in the U.S., planners find themselves working with other professionals -- lawyers, economists, and engineers. I do not feel that it can be said that planners as professionals dominate these processes or that there are institutions equivalent to conventional planning agencies. The model is much closer to the regulatory agency in which many types of professionals are engaged and around which are grouped various interests, each with its own professional support. This may be a serious problem for planning, which, in the U.S. at least, has been defined more by its institutional structure and function than by any clear set of skills and techniques. Just what is to be the contribution of planners to the process is something that schools in the U.S. are now very worried about.

Evaluation in the retrospective sense also poses challenges for planning. We have not been exempt from the criticism and examination that I

discussed earlier. Planners may expect their work to be looked at in much the same way as other public functions. As you might expect, this is not easy to do. To identify and isolate the impacts of a planning process is harder than it is for a fire department. Evaluation of planning presents formidable technical problems of measurement and control. The standard analytic model does not seem to capture what we think planning does. Yet, if we cannot identify its impacts, we may be in a very difficult position to justify our activity. It was no accident that when Proposition 13 passed in California, planning departments were among the first to be targeted for reduction by local governments.

Yet, whatever its deficiencies, planning has been able to draw upon its utopian tradition of designing a better world. Evaluation should have that same objective, but its roots are essentially analytic rather than synthetic. In a certain sense, if planning bites off too much and fails in the attempt, evaluation may bite off too little and content itself with the role of critic. Adversarial evaluation can be seen as a compromise that also reflects an approach to social design that denies the possibility of a holistic view. It says, in effect, that we must create institutions in society that let people design their own environment and resolve the differences as they go along. In architecture, this message is echoed in the advocacy of autonomous design by Christopher Alexander (1977). In the social services, it is reflected in the dissatisfaction with public helping institutions and professions, expressed in recent criticism of them as manipulative and creating the dependency that is their justification for existence [Gaylin (1978)].

I do not know what will be the outcome of all the strands that I have touched on in this discussion. To speak of an outcome as though it will be one form, itself is incorrect. We are undergoing changes in the ways in which we understand public actions and make decisions about them. I think that over the next twenty years, we may see the institutions of planning that we have known in western countries quite transformed. Evaluation seems to me an integral part of that transformation, and one that we must learn to live with and perhaps embrace.

REFERENCES

- Alexander, Christopher, Sara Ishikawa, and Murray Silverstein (1977). A Pattern Language. New York: Oxford University Press.
- Berkeley Planning Associates (1978). Evaluation of Child Abuse and Neglect Demonstration Projects. U.S. Department of Commerce, National Technical Information Service.
- Carter, Luther J. (1978). "The Attorney General and the Snail Darter," Science, 200, 628.
- Coleman, James S. (1975). "Problems of Conceptualization and Measurement in Studying Policy Impacts," in Kenneth M. Dolbeare, ed. (1975).
- Dolbeare, Kenneth M., ed. (1975). Public Policy Evaluation. Beverly Hills, Calif.: Sage Publications.
- Gaylin, Willard, Ira Glasser, Steven Marcus, and David Rothman (1978). Doing Good: The Limits of Benevolence. New York: Pantheon Books.
- Glass, Gene V., ed. (1976). Evaluation Studies Annual, Volume 1. Beverly Hills, Calif.: Sage Publications.
- Guttentag, Marcia, with Shalom Saar, eds. (1977). Evaluation Studies Review Annual, Volume 2. Beverly Hills, Calif.: Sage Publications.
- Glazer, Nathan (1978). "Should Judges Administer Social Services?" The Public Interest, 50, 64-80.
- Lynn, Lawrence E., Jr. (1977). "Policy Relevant Social Research: What Does It Look Like?" in Marcia Guttentag, ed. (1977), 63-76.
- Nagel, Theodore J. (1978). "Operating a Major Electric Utility Today," Science, 201, 985-993.
- Pressman, Jeffrey L., and Aaron B. Wildavsky (1973). Implementation. Berkeley: University of California Press.
- Teitz, Michael B. (1978). "Evaluating Program Evaluation." Journal of the American Institute of Planners, 44, 214-218.
- Webber, Melvin M. (1976). "The BART Experience: What Have We Learned?" The Public Interest, 45, 79-108.
- Weiss, Carol H. (1972). Evaluation Research. Englewood Cliffs, N.J.: Prentice-Hall.