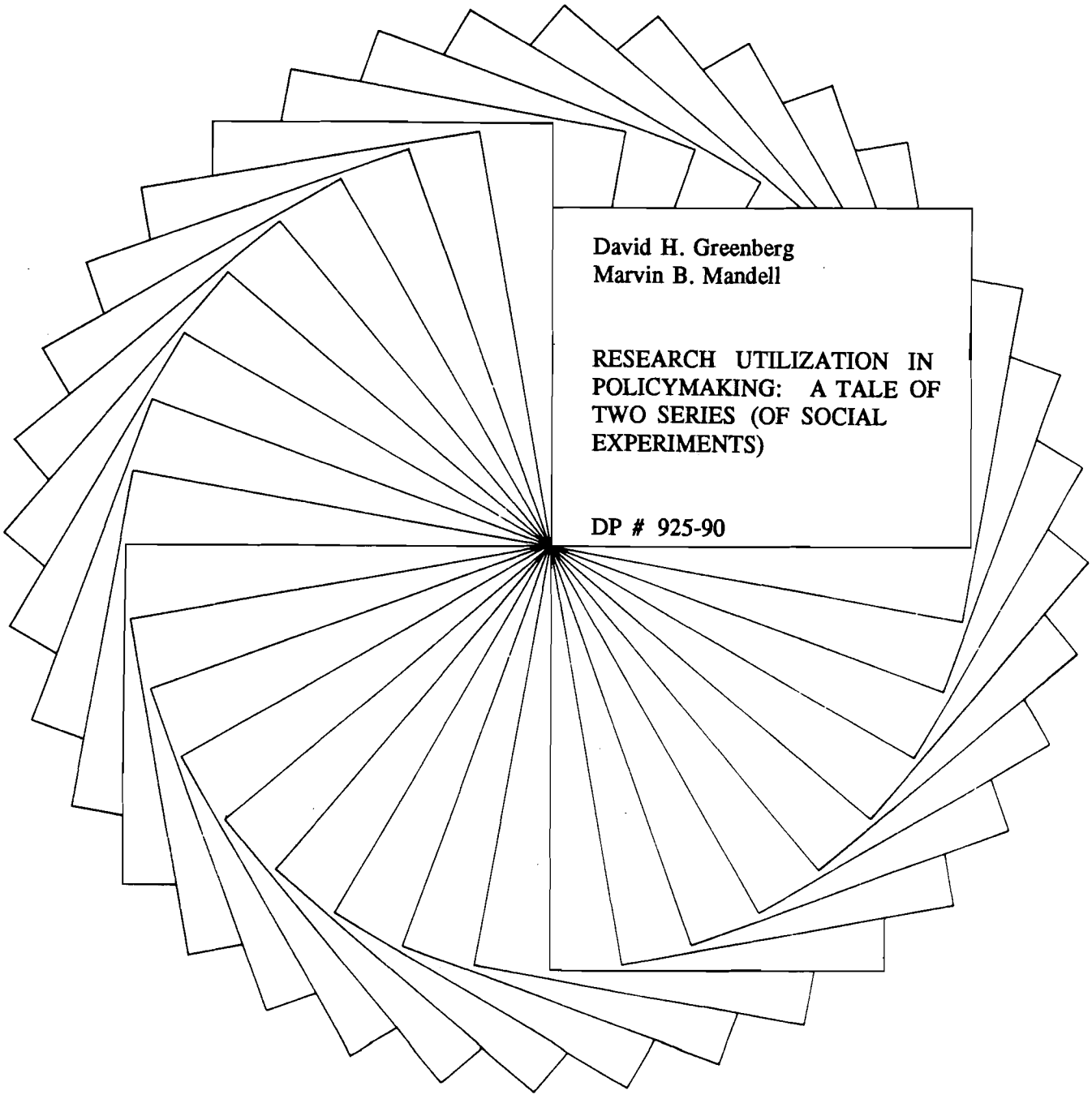
University of Wisconsin-Madison

Institute for Research on Poverty

Discussion Papers



David H. Greenberg
Marvin B. Mandell

RESEARCH UTILIZATION IN POLICYMAKING: A TALE OF TWO SERIES (OF SOCIAL EXPERIMENTS)

DP # 925-90

Institute for Research on Poverty
Discussion Paper No. 925-90

RESEARCH UTILIZATION IN POLICYMAKING:

A TALE OF TWO SERIES (OF SOCIAL EXPERIMENTS)

David H. Greenberg
and
Marvin B. Mandell
Department of Economics
University of Maryland-Baltimore County

August 1990

The research reported in this paper was supported in part by the Small Grants and Sabbatical Grants program, funded by the U.S. Department of Health and Human Services. Any opinions expressed in the paper are those of the authors alone.

Abstract

This paper represents an initial attempt to view the role that social experiments, in general, and the income maintenance experiments and the work/welfare demonstrations, in particular, have played in the policy process through the lens provided by the knowledge utilization literature. In addition to suggesting that the decision to conduct a social experiment is rarely, if ever, made according to an essentially rational paradigm, this lens helps highlight the range of uses to which findings from social experiments can be put and the circumstances under which various types of uses are more or less likely.

Preface

This paper was written as part of a larger project on social experimentation and its influence on government decisionmaking. Work on this project has been partially supported by the Small Grants and Sabbatical Grants program, which is cosponsored by the Office of Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services, and by the Institute for Research on Poverty, University of Wisconsin-Madison.

The authors of this paper are both on the faculty of the University of Maryland-Baltimore County. David Greenberg's work on the paper was performed while he was on sabbatical leave at the Institute for Research on Poverty and the Robert M. La Follette Institute of Public Affairs, both of which are located at the University of Wisconsin-Madison. The authors are responsible for the contents of this paper and the opinions and judgments expressed herein; nothing in this paper should be construed as reflecting the positions or the policy of any of the sponsoring institutions. The authors would like to acknowledge helpful comments on an earlier draft of this paper by Thomas Corbett, Robert Haveman, and Philip Robins.

While work on this paper was being conducted, a second part of the social experimentation project was also proceeding. The latter involves a comprehensive inventory of randomized social experiments in the United States. So far, over 80 different randomized social experiments have been identified. Two- to three-page summary reports are being prepared by Mark Shroder, a graduate student in the Department of Economics at the University of Wisconsin-Madison, describing each of these experiments in a uniform format. Over 60 of these summary reports, which provide information on intent, design, findings, problems, and other pertinent factors, have been completed. These summaries will be available as a Special Report published by the Institute for Research on Poverty.

RESEARCH UTILIZATION IN POLICYMAKING: A TALE OF TWO SERIES (OF SOCIAL EXPERIMENTS)

INTRODUCTION

Over the past two decades, probably over one billion dollars has been expended on conducting randomized social experiments aimed at informing policymaking (Greenberg and Robins, 1986). Yet relatively little systematic knowledge exists concerning the role such experiments play in the policy process. Why, for example, are social experiments conducted? Do they ever exert an influence on policy choices? If so, under what circumstances? The literature that does address questions of this sort is largely descriptive and atheoretical in nature. Moreover, it exists virtually independently of the "knowledge utilization" literature, which has exploded in recent years.

Overcoming the separation between these two bodies of literature can be of considerable value to both. The knowledge utilization literature contains an increasingly rich body of conceptual constructs and theoretical frameworks. Viewing the role social experiments have played in the policy process through the lenses that these concepts provide can aid considerably in interpreting and making sense of that body of experience. At the same time, the experiments represent a body of empirical data that can be valuable in overcoming weaknesses in the empirical basis previously relied upon by those in the knowledge utilization field, thereby contributing to improved testing and refinement of the conceptual constructs and theoretical frameworks developed to date in that field.

Specifically, the knowledge utilization literature has consisted largely of case studies or surveys looking at policy research in general. The former is weak in its ability to draw convincing causal conclusions. A major weakness in the latter approach is the over-generality of questions posed to respondents. For example, Webber (1986) attempted to measure the orientation of

Indiana state legislators toward the use of policy research on the basis of a four-item scale. Each item in this scale is phrased in terms of "scientific and technical information" concerning "legislative proposals," "legislative issues," or "a policy." Such a scale implicitly assumes a degree of homogeneity among "scientific and technical information," as well as among policy issues, that does not exist. Focusing on a specific set of studies--for example, social experiments--with a core of common elements can help compensate for these shortcomings in the knowledge utilization literature by enabling researchers to examine multiple cases, thereby increasing the likelihood of being able to draw strong causal conclusions, while also phrasing survey questions to respondents in ways that respect the diversity among different pieces of policy research, as well among different policy issues.

In this paper, we wish to bridge the gap between the knowledge utilization literature and the literature that addresses the role of social experiments in the policy process. Specifically, in the following sections of the paper, we look at the circumstances that make it likely a social experiment will be undertaken, the role of findings from experiments that are conducted on subsequent policy deliberations, and factors that influence whether and how findings from social experiments are utilized in making policy choices. Our discussion is illustrated by examining two sets of social experiments--the income maintenance experiments and the work/welfare demonstrations--through the lens provided by the knowledge utilization literature.

BACKGROUND

As just noted, we illustrate our discussion of the utilization of social experiments by examining the role of two specific sets of social experiments--the income maintenance experiments and the work/welfare demonstrations--in several efforts since the late 1960s to reform the nation's

welfare system. In this section, we provide background information on the particular welfare reform episodes upon which we focus and on the two sets of social experiments.

Welfare Reform Efforts

Since the mid-1960s, innumerable proposals have been made to reform the nation's welfare system. Until the late 1970s, the more prominent of these proposals embraced the concept of a negative income tax (NIT). Under this concept, the myriad of existing transfer programs available to the poor were to be collapsed into a single comprehensive program that would guarantee a minimum income payment to all households, a payment that would gradually diminish as earnings increased. A very important feature of most NIT proposals is that they would reduce, if not eliminate, the wide disparity in the treatment of different types of households under the existing welfare system. For example, two-parent households typically qualify for fewer benefits under the current welfare system than one-parent households of the same size facing similar financial circumstances. In addition, benefit levels vary widely among the states.

Despite general dissatisfaction with the existing welfare system during the 1960s and 1970s and the prominence of the NIT as the major policy alternative, a NIT was never adopted. In this paper, we shall focus on the two best known of the many welfare reform proposals that incorporated the NIT concept: the Nixon Administration's 1969 Family Assistance Plan and the Carter Administration's 1977 Program for Better Jobs and Income. The Nixon plan twice passed the House of Representatives, in 1970 and 1971, but failed both times to reach the floor of the Senate. The Carter proposal languished in committee in the House, never reaching the floor of either chamber.

For our purposes, it is not necessary to detail the specifics of either welfare proposal, except to emphasize that both contained the basic elements of a NIT. That is, both provided a

guaranteed income floor; reduced benefit as income from other sources, such as earnings, increased; replaced existing welfare programs by a single consolidated plan; and greatly reduced disparities in the treatment of different types of families by treating one- and two-parent families and families residing in different states more similarly than they are under the existing welfare system. Both plans also contained a component that was extraneous to the negative income tax concept, a work requirement. That is, able-bodied recipients, without child care responsibilities for very young children, were to be required to look for and accept employment if a suitable job was available. However, the Carter proposal went considerably further than the Nixon plan by providing minimum wage public sector jobs to eligible heads of welfare families if regular employment was unavailable.

Since the beginning of the 1980s, there has been no serious attempt to promote welfare reform proposals containing negative income tax plans as their centerpiece. Instead the thrust has been quite different, emphasizing work and reductions in transfer payment outlays and in welfare dependency. In part, this has been accomplished by tightening the requirements for receipt of welfare. In addition, the Omnibus Budget Reconciliation Act of 1981 permitted individual states to pilot test various combinations of work-oriented activities for households receiving Aid for Dependent Children (AFDC), the largest of the cash assistance programs for welfare recipients. These activities were usually intended to reduce the number of households on AFDC and were often mandatory in nature. For example, many states required participants in their AFDC Program to engage in structured job search, either individually or as part of a group. Moreover, AFDC recipients in some states were enrolled in education and training programs. And in a few states, temporary subsidies were provided to employers who hired certain welfare recipients. Finally, in some states, AFDC recipients without young children were required to work in exchange for their benefits by providing labor services to government and nonprofit agencies for a

specified number of hours per week, an activity usually referred to as "work experience" or "workfare." It was hoped that these activities would move welfare recipients into unsubsidized employment by either helping them locate appropriate jobs or providing them with work-relevant skills. In addition, it was anticipated that a work experience requirement would discourage some persons from applying for or remaining on the welfare rolls.

In marked contrast to the fate of proposals for a NIT, welfare reform legislation embodying work-oriented components has been successfully enacted. This legislation, the Family Support Act of 1988, contains several important provisions, the most pertinent of which from the perspective of this paper is the Job Opportunities and Basic Skills Training (JOBS) Program. Briefly, JOBS provides states with financial incentives to design a comprehensive work-oriented program for AFDC recipients. This program can contain various combinations of the activities mentioned above--job search, education and training, employer subsidies, and work experience. Mothers of children as young as three (or, at state option, as young as one) can be required to participate in these services. By 1995, the states will be required to keep at least 20 percent of those eligible for their JOBS programs enrolled in them. Thus, this legislation, although resulting in much more modest changes in the welfare system than would have occurred had either the Family Assistance Plan or the Program for Better Jobs and Income been enacted, shares certain critical features with the pilot plans tested under the provisions of the Omnibus Budget Reconciliation Act of 1981. For example, states were given wide latitude in designing their own work-oriented program; these programs could incorporate, at state discretion, any of the various activities tested in the earlier demonstrations; and, as in the case of the demonstrations, states could require eligible AFDC recipients to participate in these activities.

In sum then, the welfare reform debate from the mid-1960s through the late 1970s, was dominated by the NIT concept, a concept that had it been adopted would have resulted in a

massive restructuring of the welfare system. Since the late 1970s, increasing emphasis has been put on more modest, incremental reforms intended to increase the earnings and reduce the dependency of the recipient population. Each of these two very different policy concepts were tested by a separate set of random assignment experiments, the income maintenance experiments and the work/welfare demonstrations. In the next subsection, we turn to a description of these two sets of experiments.

The Experiments

Both the income maintenance experiments and the work/welfare demonstrations consist of a number of individual random assignment experiments that took place in a variety of geographic locations. These individual experiments are listed in Table 1.

The purpose of the four income maintenance experiments listed in Table 1 was to measure the magnitudes of two important effects on the behavior of the low-income population that were anticipated to result from adoption of a NIT. First, since virtually every proposed NIT program would increase the transfer payments received by two-parent households, it was expected that the members of at least some of these families would reduce the number of hours they worked. Second, because of the disparity in the treatment of one-and two-parent households under the existing welfare system, total transfer payments received by a low-income family were often higher if the father simply left. Since intact, two-parent households would be treated more similarly to one-parent households under a NIT than under the existing welfare system, it was anticipated that marital stability would be enhanced. Although the direction of these two effects was thought to be evident, considerable uncertainty existed concerning their size.

The questions raised about the work-oriented pilot programs engendered by the Omnibus Budget Reconciliation Act of 1981 were quite different from those associated with negative

Table 1

Individual Experiments Included in Income Maintenance Experiments and Work/Welfare Demonstrations

(Dates over which the demonstration treatments were administered appear in parentheses)

Income Maintenance Experiments

New Jersey/Pennsylvania Income Maintenance Experiment (8/68-9/72)

Rural Income Maintenance Experiment (rural Iowa and North Carolina) (1970-1972)

Gary (Indiana) Income Maintenance Experiment (1971-1974)

Seattle/Denver Income Maintenance Experiment (10/70-8/77)

State Work/Welfare Initiatives Demonstrations

Arkansas WORK Program (6/83-3/84)

San Diego Job Search and Work Experience Demonstration (10/82-8/83)

San Diego Saturation Demonstration (1985-1987)

Cook County (Illinois) Job Search and Work Experience Demonstration (2/85-9/85)

Maine Training Opportunities in the Private Sector Program (10/83-6/85)

Baltimore Options Program (11/82-12/83)

New Jersey On-the-Job Training Program (10/84-6/87)

Virginia Employment Services Program (8/83-9/84)

West Virginia Community Work Experience Demonstrations (7/83-4/84)

income tax programs. Among these questions were the following: Can work-oriented programs for a substantial fraction of the welfare population be administered through existing state agencies? Will the earnings of participants in these programs improve? Will these programs reduce the size of the welfare rolls? Do the benefits from these programs exceed their cost? To answer such questions, the work-oriented programs run in many states were formally evaluated.[1] The best known of these evaluations by far were of a particular subset known as the "State Work/Welfare Initiatives" demonstrations. The nine demonstrations that together comprise the work/welfare demonstrations are listed in Table 1.

The two sets of pilot tests that have just been described were similar in one critical respect: random assignment was used to allocate low-income households among two types of groups--treatment groups that were eligible for the policy changes being tested and control groups that were not. Thus, the effects of the treatments could be determined by simply comparing the treatment groups with the control groups.

The two sets of experiments differed markedly along other important dimensions, however. First, the earliest of the income maintenance experiments, the New Jersey/Pennsylvania Experiment, was initiated in 1968 and all four of these experiments were in the field by the end of 1971. The work/welfare demonstrations were initiated more than a decade later--between 1982 and 1985. Second, the income maintenance experiments were designed to test a NIT, a program that had it been adopted would have resulted in substantial structural changes in the existing welfare system. The work/welfare demonstrations, in contrast, pilot-tested various combinations of work-oriented activities that are best viewed as incrementally changing the existing system. Third, as will be discussed in more detail later, the NIT programs tested in the income maintenance experiments focused on providing income transfers to the poor and, consequently,

---

[1]For a comprehensive discussion of these evaluations, see Greenberg and Wiseman (1990).

can be characterized in terms of just a few policy parameters, an income guarantee level and a tax rate; the work-oriented programs tested in the work/welfare demonstrations, in contrast, were intended to provide services and, as a result, are structurally more complex and cannot be characterized in terms of a few parameters. Fourth, as already indicated, because the two sets of experiments tested very different policies, the outcomes of interest differed as well (behavioral effects on labor supply and marital stability versus administrative feasibility, operating costs, and impacts on earnings and welfare status). Fifth, the treatments tested in the income maintenance experiments were administered through field offices especially set up for that purpose and maintained completely separately from the existing welfare system, while the treatments tested in the work/welfare initiatives were administered through regular welfare offices. Sixth, the research design used in the income maintenance experiments was far more complex than that used in the work/welfare demonstrations. To mention just one example, none of the work/welfare demonstrations had more than one or two separate treatment groups; none of the income maintenance experiments had so few treatment groups (indeed, the Seattle/Denver Experiment alone had 47). Seventh, each of the income maintenance experiments was evaluated by a different research team, while all of the work/welfare demonstrations were evaluated by the same research organization: the Manpower Demonstration Research Corporation.

One additional dimension along which the two sets of experiments differed is in terms of their results, and in the way these results were perceived by policymakers.[2] The income maintenance experiments consistently indicated, as anticipated, that a NIT program would reduce hours of work, although the magnitude of these reductions varied among the four experiments.

---

[2]For detailed summaries of findings from the income maintenance experiments, see Munnell (1986), especially the three chapters by Burtless, Cain, and Hanushek and their discussants. Also, see Moffitt and Kehrer (1981). For detailed summaries of findings from the work/welfare demonstrations, see Burtless (1989), Greenberg and Wiseman (1990), and Gueron (1990).

Unexpectedly, however, the results did not indicate that family stability would be strengthened by a negative income tax and one of the experiments, the Seattle/Denver Experiment, seemed to imply that family dissolution would actually increase. The appropriate interpretation of both the labor supply and family stability findings for policy purposes was quite controversial at the time they initially became available and continues to remain so today (see Neuberg, 1988). For example, one can debate whether the estimated labor supply reductions are "large" or "small" or whether the dissolution of relationships that may have been tenuous to begin with is socially detrimental. Nevertheless, findings from the income maintenance experiments documented and focused attention on potentially important social costs of a negative income tax program. Thus, they placed supporters of a NIT on the defensive and generally weakened the case for adoption of such a program (see, for example, Aaron, 1978; and Burtless, 1990). As will be seen, however, this is not to say that policy decisions were necessarily based on these findings.

Findings from the work/welfare demonstrations were quite different in character from those from the income maintenance experiments. For example, they suggested that the experimental treatments were administratively feasible. Moreover, they emphasized social benefits that resulted from the tested programs. Although the treatment impacts were typically modest in magnitude and often statistically insignificant, they were consistently positive. For instance, the earnings of persons in the treatment population increased and their receipt of welfare usually decreased. In addition, there was persistent evidence of small government budgetary savings resulting from the tested treatments. There has, as yet, been little controversy concerning the proper interpretation of these results and they have been almost universally viewed as supportive of the policies they tested (see, for example, Baum, 1989; Haskins, 1990; and Szanton, 1990).

To some degree, it can be argued that one could have anticipated that findings from the work/welfare demonstrations would be more supportive of the policies they tested than would be

the case for the income maintenance experiments. After all, the income maintenance experiments estimated labor supply reductions associated with a NIT (an anticipated program cost, whatever its actual magnitude), while the work/welfare demonstrations focused on earnings improvements and reductions in welfare recipiency (anticipated program benefits). In this sense, the types of outcomes tested by the two sets of experiments may be viewed as somewhat asymmetric. However, one must be careful not to overstate this argument. Prior to the income maintenance experiments, it was generally expected that a NIT would increase marital stability (an anticipated benefit). And prior to the work/welfare demonstrations, it seemed possible that the tested programs might increase government budgetary expenditures, rather than reduce them.

WHEN ARE EXPERIMENTS CONDUCTED?

The first question of concern in examining the role in the policy process of large-scale social experiments such as those described in the previous section is when experiments are conducted. Several answers to this question might be offered. The classic answer is that the decision to conduct a social experiment is made according to an essentially rational paradigm. According to this explanation, the expected value of the information provided by a social experiment is computed and compared to the expected costs of conducting that experiment. The relevant procedures are well documented in the decision analysis literature (e.g., Thompson, 1982). This model would predict that experiments are conducted when the expected value of the information provided by a social experiment exceeds the expected costs of conducting that experiment.

While appealing, the inconsistency between the requirements of this model, on the one hand, and the characteristics of social experiments and the policy processes in which they are embedded, on the other hand, poses significant challenges to the explanation that the decision to conduct a social experiment follows an essentially rational paradigm. Specifically, Feldman

(1989: 80) contends that to be considered consistent with the standards set by a rational or bounded rational model, an analysis should be intended for a specific policy decision. In addition, according to a rational model, the time frame of this policy decision, which is assumed to be known, dictates the timing of the analysis in that the latter is completed by the time the policy decision is to be made. This model further requires that at the time a choice is made as to whether or not to conduct a social experiment, the alternatives that will be considered in the policy process, along with an a priori probability distribution for the consequences of each alternative, must be known. Moreover, in order to make these calculations, decisionmakers' preference functions over these consequences must also be specified at the time the choice concerning conducting the social experiment is made. This, in turn, presupposes that those who will participate in the policy decision are identified at the time the choice on conducting the social experiment is made.

In short, the explanation that the decision as to whether or not to conduct a social experiment follows an essential rational paradigm requires a policy environment that is highly predictable and orderly. In fact, however, the policy environment generally associated with social experiments is highly fluid and volatile. Such matters as what issues will be on the agenda, what forms those issues will take, what alternatives will be seriously considered, and who will be participating in the relevant choice processes are extremely unpredictable (see, for example, Kingdon, 1984; March and Olsen, 1976; Weiss, 1980). The difficulty of predicting these sorts of matters casts doubts upon the adequacy of a rational paradigm for making predictions concerning the conduct of any type of policy research, but is especially problematic in the case of social experiments since such experiments typically have an especially long time frame. For example, the income maintenance experiments were initiated in 1968. Had the intent been to use them in consideration of the Nixon Administration's Family Assistance Plan, they would have been

completed by the time that plan was being considered by Congress in 1970 and 1971. In fact, however, the only results from the income maintenance experiments that were available at that time were some highly preliminary results from the New Jersey/Pennsylvania Experiment--and analysts released these only under considerable duress. Thus, it appears highly unlikely that the income maintenance experiments were intended for the 1970-71 consideration of the Family Assistance Plan. It seems even less likely that the decision to conduct these experiments was influenced by anticipation of consideration of welfare reform in the Carter Administration, a decade later.

Similarly, as Baum (1989) notes, at the time the work/welfare demonstrations were initiated in 1982, the prospects for welfare reform looked especially bleak. She argues that a number of events in the policy environment between 1982 and 1988, including the 1986 elections (which resulted in the Senate reverting from Republican to Democratic control and, consequently, made Senator Moynihan Chair of the Subcommittee on Social Security and Family Policy of the Senate Committee on Finance in the 100th Congress) were necessary for welfare reform to be put onto the agenda. It seems highly unlikely that events such as these could have been predicted by anybody in 1982.

If the rational paradigm does not offer an adequate explanation for the decision to conduct social experiments, what other explanations might be offered? Feldman (1989: 92-96), in examining bureaucratic policy analysis in general, argues that the unpredictability of the policy environment makes it impossible to specify when and for what issues policymakers will seek research results. She suggests that policy research be viewed in terms of creating inventories of information for future policymaking situations that are unspecified at the time the information-creation process begins. Applying this argument to experiments, a social experiment can be seen as simply adding another item to the information inventory--one that should help reduce

uncertainty concerning alternatives--just in case a pertinent policy issue is actively considered (for example, welfare reform in the context of the income maintenance or the work/welfare demonstrations).

While not based on the decision-theoretic calculations of the rational paradigm, this explanation, it should be noted, still focuses on the value of the products of the experiment. The difficulty with this argument when applied to social experiments is the high costs of many social experiments. It seems implausible that analysts or policymakers would be willing or able to invest such large amounts of resources in an activity that has only a limited probability of ever being called upon for guidance.

Another set of possible explanations of the decision to conduct a social experiment focuses not on the anticipated future products of the experiment, but on benefits from the process of conducting the experiment. These benefits, it should be noted, begin to accrue virtually as soon as the process of designing and conducting the experiment is started. A traditional version of this set of explanations would say something to the effect that experiments are conducted to replace real action with symbolic action, thereby serving as an inherently conservative force.

Perhaps a more interesting variant of this type of explanation is that experiments are a means of keeping alive innovative policy alternatives for which adequate support cannot be obtained immediately. As compared to the more traditional approach of this argument, this version sees social experiments serving as a (modest) force for change. It suggests that in the absence of social experiments, whether the innovative policy alternatives ever appear on a future political agenda will be a purely random event. Conducting social experiments provides an extra impetus for those alternatives appearing on the agenda at the time the results become available, at least if the findings are generally supportive of the tested social innovation. Another implication of this argument is that it suggests that political negotiation will proceed in parallel

with the process of designing and conducting a social experiment. Hence, a political agreement might very well be reached before the results of the experiment are available, in which case any impact those results have on policy deliberations would be "icing on the cake," but not really essential. There is reason to believe that this explanation accounts, in part, for the conduct of both the income maintenance experiments and the work/welfare demonstrations.

POLICY EFFECTS OF THE EXPERIMENTS

The work/welfare demonstrations tested a set of policies that became incorporated into law as part of the Family Support Act of 1988, while the income maintenance experiments tested a program that provided the centerpiece of several legislative proposals that failed, including the Family Assistance Plan and the Program for Better Jobs and Income. By itself, however, this does not demonstrate that either set of social experiments had an effect upon policymaking–that is, that findings from the work/welfare demonstrations resulted in adoption of the JOBS Program or that findings from the income maintenance experiments precluded enactment of a negative income tax program. Indeed, as Haveman (1987: 198) points out, no definitive conclusion concerning the policy effects of social experiments is possible without first determining "the policy impacts of the foregone activities--both research and nonresearch--which would have occurred but never did." Since this is virtually impossible to determine, we do not ask the standard counterfactual question: How would policy have differed had the experiments never been conducted? Instead, we ask a different question, one that is somewhat easier to answer, although a definitive answer is probably still not possible: How would policy have differed had findings from the experiments been radically different from what they actually were? Specifically, what would have happened had the income maintenance experiments indicated that labor supply would be unaffected and family stability strengthened by adoption of a NIT? Or had the work/welfare

demonstrations implied that the tested treatments were administratively unmanageable, had no effect on earnings and welfare status, and increased government budgetary cost?

In considering these questions, two contributions made by the "knowledge utilization" literature are especially relevant. The first is made by that portion of the "knowledge utilization" concerned with clarifying the concept of research utilization and identifying different types of utilization. The second is made by Weiss's (1983) I-I-I (Ideology-Interest-Information) framework.

A major theme in the former stream of literature is that the conventional conceptualization of utilization, in which the results of a specific study (or set of studies) decisively affect a pending decision, is only one of many types of utilization. Types of utilization other than instrumental utilization, as the conventional conceptualization is known, tend to be less dramatic, far more subtle and, consequently, more difficult to detect than instrumental utilization.[3]

Whiteman's (1985) two-dimensional framework, which is shown in Table 2, integrates most of the various types of utilization that have been described in the literature. While Whiteman collapses each dimension into discrete categories for convenience, each is, in fact, a continuum. The first dimension is referred to as the concrete/conceptual dimension. Concrete utilization refers to the direct effects of specific conclusions and findings from policy research on specific decisions. Conceptual utilization, in contrast, refers to the effects of research on the general

[3]Different writers have called the utilization model that applies to instrumental use "the rational model," "the decision model," "the decision-driven model," "the decisionistic-instrumental model," and "the problem-solving model." Although these terms are not used to mean exactly the same thing, the overlap is considerable.

Table 2

Uses of Income Maintenance Experiments and Work/Welfare Demonstrations

	Substantive	Elaborative	Strategic	
			Symbolic	Persuasive
Concrete		Both influenced program design.	Both were used to justify, explain, and advocate positions.	Only the work/welfare demonstrations strengthened the hand of those who used it to justify, explain, and advocate positions.
		Income maintenance experiments influenced adoption of retrospective accounting and monthly reporting.		
Conceptual	Income maintenance experiments helped persuade some long time supporters of a negative income tax to relinquish that position.	Income maintenance experiments refined understanding of labor supply elasticity.		Income maintenance experiments confirmed understanding of labor supply elasticity.

intellectual orientations of policy actors. Conceptual utilization might, of course, affect specific decisions in the future, but in an indirect way.[4]

The second dimension in Whiteman's framework is the substantive/elaborative/strategic dimension. Substantive use occurs when research determines the basic outlines of an individual's position (concrete-substantive use) or orientation (conceptual-substantive). Elaborative use is said to occur when research refines a policy actor's position or general orientation "within basic outlines which have already been established" (Whiteman, 1985: 206). Strategic use refers to the role of policy research in "advocat[ing] and reaffirm[ing] policy positions after they have been determined" (Whiteman, 1985: 206). Thus, the major distinction between substantive and strategic use is that in the former policy research forms part of a policy actor's predecision argument, while in the latter policy research forms part of a policy actor's postdecision argument used for justification, explanation, and persuasion. (For a more in-depth discussion of the importance and value of postdecision arguments, see Majone, 1989: chapter 2.)

It is important to note, that just as conceptual utilization can lead indirectly to substantive use, so might strategic use affect specific decisions. That is, strategic use often enables policy actors to present a more coherent and forceful argument than they would otherwise be able to present. This, in turn, might make these actors more persuasive or influential as the policy deliberations proceed, thereby affecting the outcome of those deliberations. This suggests the possibility that the category of uses which Whiteman (1985) labels strategic, in fact, contains two

---

[4]There are a number of overlapping models that apply to conceptual uses of research including "the interactive model" (Weiss, 1979), "the limestone model" (Thomas, 1987), "the enlightment model" (Weiss, 1979), and "the guano model" (Vining, as quoted by Hanushek and Weimer, 1989). To illustrate: the limestone model portrays the influence of social science research as analogous to "the action of water through limestone. You may know where the water falls on the limestone, but there is no means of knowing what route it will take down the various levels, or where it will emerge through unexpected fissures" (Robert Guthrie, as quoted by Thomas). The guano model implies that each piece of research makes a small contribution by itself; but, just like bird droppings, together it all adds up.

distinguishable subcategories. The first of these sub-categories, which we label "symbolic," includes those situations in which policy research is used to advocate and reaffirm policy positions after these positions are already determined. The second subcategory, which we label "persuasive," is distinguished from symbolic use in that referring to policy research in the course of advocating and reaffirming policy positions actually affects the outcomes of the policy deliberations.

In contrast to the original expectations of those sponsoring and conducting the income maintenance experiments and the work/welfare demonstrations, neither of these sets of experiments resulted in concrete-substantive use. As previously mentioned, with the exception of some highly preliminary results from the New Jersey/Pennsylvania Experiment, findings from the income maintenance experiments were not available at the time the Family Assistance Plan was being considered by Congress in 1970 and 1971. More important, the income maintenance experiments were designed to test the effects of NIT programs on labor supply and marital stability, but the fates of both the Family Assistance Plan and the Program for Better Jobs and Income were not decided solely on the basis of such behavioral effects. Other issues were at least as important.

Detailed discussions of these issues may be found elsewhere (see, for example, Moynihan, 1973; Burke and Burke, 1974; Coyle and Wildavsky, 1986; Aaron, 1973, 1978; 1984; Lynn and Whitman, 1981) and, consequently, we simply list a sampling. First, and most fundamentally, many persons strongly objected philosophically to the central concept of a NIT: providing a guaranteed income for persons capable of working. As Aaron (1990:277) has written "the public and their elected representatives have made clear in every way they know how that they think it is good for healthy working-age people to receive pay in return for work and bad for them to get pay without work." Second, there was concern over the additional budgetary costs that would be engendered

by enactment of a NIT program and the number of able-bodied adults who would be added to the welfare rolls. Third, there was resistance in low payment states to increasing the benefit amounts of welfare recipients in these states. Fourth, considerable controversy occurred over the work test incorporated into the Family Assistance Plan and the public jobs component of the Program for Better Jobs and Income.

The absence of concrete-substantive use in the NIT case is consistent with Weiss's I-I-I framework. The central premise of the I-I-I model is that "the policy positions taken by policy actors are the resultant of...the interplay of ideology, interests, and information..." (Weiss, 1983: 221). Ideology, as Weiss uses it, "encompasses...principles, values, [and] political orientation. At its core are ethical and moral values" (Weiss, 1983: 224). She defines interests "primarily in terms of self-interest" (Weiss, 1983: 224). Information is descriptive in nature, and is defined as policy actors' "sense of the current state of affairs, the relative seriousness of problems, [and] why things happen as they do..."(Weiss, 1983: 225). Weiss notes that policy research is only one source of information; other sources include direct experience, craft lore, media reports, formal and informal organizational channels, and interest groups. An important hypothesis that Weiss (1983: 243) derives from the I-I-I framework is that the greater the internal consistency among ideologies, interests and information from sources other than research, the less the utilization of research results.[5] At least with respect to concrete-substantive use, the experience in the NIT case is consistent with this hypothesis.

---

[5]A somewhat different way of looking at this is suggested by Jenkins-Smith (1990: 172), who writes that "it is in the midrange of conflict--where some incentive exists to mobilize resources for analysis yet change in belief systems would be tolerable--that the results of [policy] analysis could be expected to affect the beliefs of policy elites. Low-level conflict would be expected to draw little analytical attention, and when conflict is high, analysis will tend to be deployed primarily as a political tool to promote, justify and defend a predetermined policy position." The level of conflict over the policies tested in the income maintenance experiments was obviously much higher than that over the policies tested in the work/welfare demonstrations.

Although the income maintenance experiments appear not to have resulted in concrete-substantive use, they did result in other forms of concrete use. In particular, they had considerable concrete-elaborative use. As Reischauer (1986: 215) has pointed out, "the findings from the [income maintenance] experiments were discussed at dozens of planning and strategy meetings between congressional advocates of welfare reform and policy formulators in the executive branch. They influenced the design of the Carter welfare reform plan in numerous ways."

In addition, it is obvious that the income maintenance experiments resulted in concrete-symbolic uses in that they provided ammunition for those who argued against the Program for Better Jobs and Income. However, it is doubtful that these experiments had a concrete-persuasive effect. While the results of the income maintenance experiments were cited by opponents of the Program for Better Jobs and Income, having such ammunition does not appear to have significantly affected the outcomes of the policy deliberations by making opponents more persuasive or influential than they would have would been in the absence of this ammunition. It simply tended to reinforce the positions of those who were against NITs on other grounds, while failing to dislodge those who already favored NITs. In the same vein, had supportive findings (that is findings to the effect that labor supply would be unaffected and family stability strengthened by adoption of a NIT) been obtained from the experiments, it is virtually certain that proponents of a NIT would have used the results to justify, explain, and advocate their position. But in light of the strengths with which positions were held due to existing ideologies, information from other sources, and political interests, it seems unlikely that this would have altered the outcomes of those deliberations by making the proponents of NITs sufficiently more persuasive or influential.

In summary, then, the income maintenance experiments suggest that when there is a high degree of internal consistency among ideologies, interests, and information from sources other than research, concrete-substantive and concrete-persuasive uses are very unlikely. However, a high degree of I-I-I consistency does not appear to reduce greatly the likelihood of concrete-elaborative or concrete-symbolic use.

Like the income maintenance experiments, the work/welfare demonstrations did not result in concrete-substantive use in that the major thrusts of the Family Support Act of 1988 were specified prior to the availability of MDRC's work (Baum, 1989). However, findings from the work/welfare demonstrations appear to have had concrete-elaborative use in that they provided input into the details of the Family Support Act of 1988. Baum (1989: 22), who worked for Senator Moynihan in drafting the bill, seems to have captured this quite well when she states that by the time findings from the work/welfare demonstrations were taken into account, "most of the key provisions of the JOBS program...had already been worked out....[but] we used [the findings] to modify certain provisions."

Moreover, findings from the work/welfare demonstrations resulted in concrete-symbolic use by helping provide ammunition to those who favored the legislation. Unlike the income maintenance experiments, however, concrete-symbolic use was in this case accompanied by concrete-persuasive use. As both Baum (1989) and Szanton (1990) amply document, while the major thrusts of the Family Support Act of 1988 were specified prior to the availability of MDRC's work, passage of welfare reform legislation in 1988 was far from a foregone conclusion. Considerable uncertainty existed over whether work-oriented programs "worked" for the welfare population. The results of the work/welfare demonstrations greatly strengthened the ability of the proponents of welfare reform to formulate a persuasive postdecision argument. As Baum notes "Later, when the bill was formally introduced,...[they] also gave us powerful ammunition to use in

fending off critics from both ends of the political spectrum." Szanton (1990) quotes several congressional staffers to this effect. To cite just one example (p. 11): "'In all the years I worked on welfare reform, we never had a body of data that showed what worked. Now we had it.'" It is therefore quite plausible that the work-oriented policies embodied in the Family Support Act would not have been enacted into law had the work/welfare demonstrations been perceived as demonstrating that these policies were ineffective.

Nonetheless, as Weiss's I-I-I framework suggests, there were limits on the extent to which the work/welfare demonstrations could have concrete-substantive and concrete-persuasive effects, even though the configuration of ideology, interests, and existing information had much less internal consistency than was the case in the NIT episodes. Baum (1989: 29) provides two examples of these limitations: the amendment to the legislation that required 16 hours of community work experience for two-parent AFDC families and the amendment that added minimum participation rate requirements. On both of these issues, Baum claims, "key players had already made up their minds -- on the basis of ...[among other things] political ideology." On these two issues, Baum notes, "new, reliable research data seemed to have very little effect."

Another important use of the negative income experiments concerns retrospective accounting and monthly reporting. This is an interesting example because it is somewhere in the middle of both dimensions in the framework. As a result of being used in administering the NIT plans tested in the income maintenance experiments, these two related innovations appeared to be quite promising. Following the income maintenance experiments, a series of subsequent experiments tested them explicitly. Early results from these follow-up experiments indicated that retrospective accounting and monthly reporting were effective in reducing welfare expenditures, and these innovations were then adopted nationwide in administering the existing welfare system.

It would be inaccurate to label this episode a concrete use of the income maintenance experiments, since the linkage from the income maintenance experiments to the adoption of these innovations is not a direct one. Yet, it would also be inaccurate to label this a conceptual use of the income maintenance experiments, since the story ends in a specific decision. Moreover, inasmuch as policymakers were not already convinced of the merits of retrospective accounting and monthly reporting prior to the income maintenance experiments, this does not represent a strategic use. But neither does it represent substantive use, since the introduction of retrospective accounting and monthly reporting did not change the core of welfare policy. Rather, this is another illustration of elaborative use.

Perhaps more important than the concrete uses of the income maintenance experiments are their conceptual uses. Somewhat conjecturally, findings from the income maintenance experiments may have had conceptual-substantive effects in that they may have helped motivate some long time supporters of a negative income tax to relinquish this position and, consequently, helped pave the way for the very different policies embodied in the Family Support Act. (To the extent that this is true, it illustrates the earlier point that conceptual utilization might indirectly affect specific decisions in the future.) For example, Daniel Patrick Moynihan, who had generally been supportive of the NIT concept, first as an official in the Nixon administration and later as a U.S. Senator, appeared, as a consequence of the family stability findings from the Seattle-Denver Experiment, to have had a change of heart in the late 1970s (Elmore, 1986: 209). In 1988, Moynihan became the major sponsor of the Family Support Act in the Senate. Not too much should be made of Moynihan's transition, however. Erosion of support for the NIT probably had more to do with the past political failures of this concept, the growing budget deficit, and the ushering in of a conservative political period heralded by the election of President Reagan in 1980 than with findings from the income maintenance experiments.

The income maintenance experiments appear also to have had considerable conceptual-elaborative and conceptual-persuasive effects, especially by subtly altering attitudes over the long-run. For example, prior to the income maintenance experiments, supporters of guaranteed income programs, such as NITs, tended to dismiss potential program effects on labor supply as unimportant, while those opposed to such programs often claimed that they would cause massive withdrawals from the work force. Today, either extreme position is no longer tenable in the face of findings from these experiments.

More fundamentally, the income maintenance experiments refined and confirmed understanding of labor supply elasticity. For example, findings from the income maintenance experiments strongly implied that the labor supply of prime-age males is quite inelastic, while the labor supply of married females is considerably more elastic. Similar results had previously been obtained from various cross-sectional studies, and thus, the income maintenance experiments did not radically change beliefs concerning labor supply elasticity. But the income maintenance experiments provided more precise estimates of the magnitude of these elasticities than had previously been available, thereby reducing the range of uncertainty (Stafford, 1985). As a result of the income maintenance experiments, labor economists and policymakers could be more confident concerning magnitudes of labor supply elasticities for various groups than they previously had been. These magnitudes are, in turn, a pertinent concern in a variety of policy issues including virtually all changes in tax and transfer programs, as well as policies that directly intervene in the labor market--for example, training programs and the minimum wage.

In contrast to the income maintenance experiments, conceptual uses of the work/welfare demonstrations appear to have been minimal. In part, this might simply be due to the relatively short amount of time that has elapsed since the results of this set of experiments first became available. But our speculation is that there is more to it than this. As we discuss in more detail

below, the income maintenance experiments may be characterized as "response surface" experiments and the work/welfare demonstrations as "black box" experiments (see Orr, 1985). It is plausible that the greater generalizability of the former type of experiment, as compared to the latter, significantly increases the prospects that such experiments will result in conceptual uses.

UTILIZATION FACTORS

The literature on research utilization suggests specific factors that can be used in ranking research in terms of its probable utilization. Although the exact taxonomy differs somewhat from author to author, the five following factors appear to capture the essence of these analyses: credibility, timelineness, communication and visibility, generalizability, and relevance.

Each of these factors are discussed in some detail below and their relevance to the income maintenance and the work/welfare experiments is described.

Credibility

Interestingly, empirical findings on the relationship between the scientific quality of research and utilization are inconsistent and conflicting (see Beyer and Trice, 1982: 610; Mandell and Sauter, 1984). Nevertheless, there appears to be little controversy over the proposition that research is likely to exert more influence on policymakers if it is perceived as valid by them than if it is not. For any particular piece of research, a number of different considerations have been found to affect credibility: whether there is debate among researchers concerning the validity of the findings (such debate tends to demystify the technical expertise of researchers) (Sabatier, 1978: 409); whether the findings are ambiguous; the extent to which the findings corroborate or contradict other information sources or the intuition, values, experience, and expectations of potential users (Leviton and Hughes, 1981: 539; Weiss and Bucuvalas, 1980); the extent to which

the findings are threatening to potential users' goals or imply that changes in potential users' current activities are needed (Sabatier, 1978: 410, suggests that this is because, as implied by cognitive dissonance theory, "information that does not conform to policy predispositions will be ignored or distorted by the decision maker"); the opinions of potential users towards research (Leviton and Hughes, 1981: 540 note that there is evidence, for example, that doctors are more receptive to research than lawyers; also see Bozeman, 1986: 531); characteristics of those who produced the research (Bozeman, 1986: 531; Coursey, in progress; Leviton and Hughes, 1981: 540; Mandell, 1989: 45); and the extent to which the research is subject to scrutinization (that is, to attacks on the methodological quality of the research).

At first blush, it might appear that research findings that are based on a random assignment social experiment would be more credible among policymakers than findings based on non-experimental techniques. After all, an experimental research design is much less vulnerable to methodological criticism than alternative designs and is also readily understood by policymakers. Indeed, John Ziman, a physicist, has suggested that "a good experiment is a powerful piece of rhetoric; it has the ability to persuade the most obdurate and skeptical mind to accept a new idea..."(quoted by Majone, 1989: 30-31). Nevertheless, consideration of the factors listed in the preceding paragraph suggests that social experiments need not necessarily be more credible than nonexperimental research. One important reason for this is that the very size and importance of a large-scale series of social experiments tends to attract both scrutinization and reanalysis of the experimental data. Moreover, experimental findings are not necessarily less likely than nonexperimental findings to contradict expectations or be ambiguous in nature.

These points are well illustrated by experience with the income maintenance experiments during the Carter administration's attempts to reform the welfare system in the late 1970s.

Although results from these experiments received considerable publicity during this period, several circumstances tended to undermine their credibility.

First, a single, internally consistent story did not emerge from the four income maintenance experiments. Although the largest, most expensive, and most recent of these experiments--the Seattle-Denver Experiment--received the greatest attention, the fact remains that a multitude of varying results were available. It is true that all four experiments were consistent in implying that the experimental treatments tended to reduce hours of work. However, the estimated magnitude of this effect differed considerably across the experiments. Moreover, of the four experiments, only data from the Seattle-Denver Experiment seemed to provide strong evidence that marital stability would be adversely effected by the treatment. Possibly more importantly, since analyses of the experimental data were conducted by four separate research teams, the statistical techniques used and the formats in which findings were reported varied. In addition, there were several alternative sources of information about the experimental results, each of which tended to interpret the findings somewhat differently. These sources included analysts who were part of the original four research teams; analysts at the Department of Health and Human Services' Office of the Assistant Secretary of Planning and Evaluation, the funding agency for the experiments; and increasingly, analysts at other government agencies and at the universities.

Second, the experiments were subject to considerable scrutiny by outside researchers, as, indeed, might have been anticipated given the level of controversy over the NIT concept and the fact that the income maintenance experiments were the first set of large-scale social experiments ever conducted. Well before the Carter administration's welfare reform efforts began, the Brookings Institution held separate conferences on the New Jersey and the Rural income maintenance experiments, the first two of the four experiments. The two experiments were thoroughly critiqued in papers presented at these conferences, which were later published

(Pechman and Timpane, 1975; Palmer and Pechman, 1978). An additional book length critique of the New Jersey experiment was sponsored by the Russell Sage Foundation (Rossi and Lyall, 1976). Moreover, toward the end of the 1970s, data from the income maintenance experiments began to be reanalyzed by economists who focused on certain methodological issues associated with the labor supply findings and purported to demonstrate that the initially estimated treatment effects on hours of work were either seriously understated or overstated (see, for example, Cogan, 1978; Greenberg, Moffitt, and Friedman, 1981; and Burtless and Greenberg, 1982). Although the last two of these studies were not published in academic journals until the early 1980s, findings from all three studies were available sufficiently early to begin to muddy the waters while welfare reform was still under active consideration.

Third, some of the findings from the income maintenance experiments were counter-intuitive and unexpected. In particular, as mentioned earlier, it was anticipated that, relative to the existing welfare system, the experimental treatments would strengthen marital stability since these treatments appeared to provide much less of a financial incentive to separate. Consequently, the apparent adverse effect of Seattle-Denver treatment on marital stability was almost entirely unexpected by the policymaking community. Thus, these findings were instantaneously controversial. Some persons used them symbolically to argue against the welfare reforms being considered by the Carter administration. Others, mainly supporters of welfare reform, attempted to dismiss the findings by suggesting statistical reasons for not taking them seriously or arguing that if the experimental payments allowed women to get out of an unhappy marriage, this effect was, at worst, benign.

While numerous factors tended to undermine the credibility of the income maintenance experiments during the Carter administration, the work/welfare experiments seemed to enjoy enormous credibility when welfare reform was again being seriously considered about a decade

later. This was not by accident. In marked contrast to the income maintenance experiments, all the work/welfare demonstrations that were pertinent to policy considerations during the late 1980s were evaluated by a single organization, the Manpower Demonstration Research Corporation (MDRC), that used a virtually identical methodology in all its analyses and followed a similar format in reporting its findings.[6] Moreover, these findings were interpreted almost exclusively by a single very articulate spokesperson: Judith M. Gueron, the president of MDRC. Furthermore, while the different work/welfare demonstrations did not produce entirely uniform results, the findings did seem to fit into a consistent pattern. It appeared that, in general, the experimental treatments modestly increased the unsubsidized earnings of welfare recipients and decreased transfer program expenditures. These outcomes were viewed by almost everyone as positive, consistent with expectations, and noncontroversial. In addition, the work/welfare demonstrations were administered by existing state agencies, giving them an aura of realism. The income maintenance experiments, in contrast, were administered by field operations specifically establish and run by government contractors for the purpose. Finally, MDRC's evaluations of the work/welfare demonstrations were subjected to little scrutiny by outside researchers prior to and during the time the 1988 Family Support Act was under consideration, and no reanalysis of data from these demonstrations had yet been undertaken. One reason for this is that, unlike the research organizations that evaluated the income maintenance experiments, MDRC was neither provided funding nor contractually obligated to make data from the work/welfare demonstrations available to outside researchers and, indeed, until very recently, had not.

---

[6]Although there were a number of evaluations of work/welfare demonstrations in which MDRC did not participate (see Greenberg and Wiseman, 1990, for a description), only a few of these were based on experimental evaluation designs and, more importantly, results from these evaluations were virtually unknown outside of the states in which the demonstrations were conducted.

Timeliness

Timeliness implies that research utilization depends upon the research reaching decisionmakers when, to use John Kingdon's (1984) term, "policy windows" are open (that is, when decisions are actually being made). Timeliness is obviously more important for concrete uses than for conceptual uses (Leviton and Hughes, 1981: 525). In stressing the importance of timeliness in congressional decisionmaking, Weiss (1987: 106) has suggested that to be used "policy analysis probably needs to get into the pipeline while general approaches to legislation are still being considered....By the time Congressional committees begin action on specific bills, the determining conceptional premises have usually started to set." Similarly, James Coleman (1979) has argued that the greatest source of incompatibility between research and policy may be the fact that policy decisions have time schedules that are often in serious conflict with the time schedules of research.

In part, the importance of timeliness depends on how long "policy windows" remain open or how often they reopen. It also depends on whether decisionmakers tend to utilize inventories of previously conducted research (Feldman, 1985), or only newly completed research. This, in turn, is partially determined by whether findings from the older research continue to be perceived as relevant to current policy concerns, an issue we take up later.

Relative to other research, social experiments would appear to be at a particular disadvantage regarding timeliness. By the time the experiment is designed and implemented, the treatment administered, and the data analyzed, several years can easily transpire. To mention an extreme example, the Seattle-Denver Income Maintenance Experiment was initiated in 1970, preliminary results began to be available in 1976, and the final report was not published until

1983.[7] Perhaps more typically, the San Diego Job Search and Work Experience Demonstration, one of the earliest and more prominent of the work/welfare demonstrations evaluated by MDRC, began in 1982 and the final report was published in February 1986.

Problems caused by attempts to circumvent these lengthy time lags are illustrated by a controversial incident that occurred in early 1970 when the Nixon Administration was promoting its welfare reform proposal. In an attempt to provide experimental evidence in time to contribute to debate over this issue, an analysis based on preliminary and incomplete data from the New Jersey Income Maintenance Experiment was hastily conducted. Although findings from this analysis suggested that the experimental treatment did not cause work disincentives and, consequently, seemed to support the Nixon proposal, they were soon declared "premature" by the General Accounting Office, which indeed they were. As mentioned earlier, treatment under the income maintenance experiments was ultimately found to reduce hours of work.[8]

Although the possibility of attempting to consume the fruit from a social experiment before it is ripe certainly exists, problems of timeliness associated with social experimentation may not be as severe as sometimes supposed. In fact, findings from all four income maintenance experiments were available in time for debate over the welfare reform efforts of the late 1970s and final reports on almost all the work/welfare demonstrations were published prior to legislative consideration of the 1988 Family Support Act. Although this may appear purely serendipitous, this is not necessarily the case. As suggested earlier, social experiments may exert some influence on the political agenda. And sometimes policy windows reopen. For example, as pointed out

---

[7]Indeed, important findings from this experiment continued to appear even after 1983. For example, a major reanalysis of the effects of the Seattle-Denver treatment on marital stability, one that purported to overturn the findings appearing in the final report, was not published until two decades after work on the experiment first began (Cain and Wissoker, 1990).

[8]For greater detail concerning this incident, see Coyle and Wildavsky (1986).

earlier, there were important similarities in the Nixon and Carter administration welfare reform proposals, and findings from the income maintenance experiments were applicable to both. Moreover, some policy concepts germinate slowly. The policies embodied in the Family Support Act evolved over a lengthy period of time and some of these policies were tested by the work/welfare demonstrations. Indeed, as previously argued, had these tests not been perceived as successful, the legislation may have been written differently or might not have become law. Thus, under certain circumstances, a social experiment may help open policy windows or, at least, alter the view seen through these windows.

Communication and Visibility

It seems self-evident that research utilization is more likely the more aware decisionmakers are of findings from research and the better they understand them. Social experiments would appear to enjoy an important inherent advantage over other types of social science research in this respect since experimental findings can be reported in terms of simple comparisons of differences between experimentals and controls, statistics that are readily communicated to and understood by non-technicians (see Mundel, 1985: 254). Moreover, the larger-scale social experiments are highly visible. Aaron (1985: 276) suggests that "the fact that social experiments are political events would appear to give them a great advantage" in this respect. Indeed, findings from social experiments have been frequently reported in local newspapers.

The literature on utilization, however, suggests that communication and visibility does not depend just on characteristics inherent in the type of research being conducted. It is also influenced by dissemination efforts on the part of those conducting the research and may not occur unless researchers take the initiative (Beyer and Trice, 1982: 602; Mandell, 1988). Even if these efforts are undertaken, however, their effectiveness depends upon presentation--for

example, length (decisionmakers face severe time and energy constraints), whether written or oral (many decisionmakers rarely read), clarity, and whether jargon is used (Cox, 1977; Mintzberg, 1971; Sproull and Larkey, 1979; Weiss, 1984). Dissemination may also be facilitated by the existence of what Heclo (1978) has called an "issue network." Such a network of researchers and policy analysts, both within and outside the government, has long existed around the issue of welfare reform.

Just as the dissemination effort on the part of researchers may affect utilization, so may advocacy on the user side. The literature suggests that advocacy of the research results on the part of a key individual or two may be crucial to concrete use of the findings (Leviton and Hughes, 1981: 541-542 and Weiss, 1984). Thus, it may be important, as Patton and his associates (1977: 158) have suggested, that someone within the organization that can potentially make use of the research takes "direct, personal responsibility for getting information to the right people...." It is also sometimes suggested that utilization may be enhanced if potential users participate in the research (Beyer and Trice, 1982: 607; Bryk, 1983; Mandell, 1984), as occurs in the case of a social experiment run through existing administrative units. Advocacy, of course, is only likely if the findings help promote the key individuals' goals.

Since they were large and among the first of the social experiments, the income maintenance experiments were highly publicized and well known. Moreover, numerous reports were published, executive summaries of findings were circulated, and congressional staff were briefed. And on at least two occasions, formal testimony was given before congressional committees concerning results from the experiments (see Elmore, 1986: 209). As already mentioned, however, information about the income maintenance experiments was provided by a substantial number of different information sources and did not seem to provide a single coherent picture. Moreover, although results were sometimes presented in the form of simple differences between

experimentals and controls, it was frequently the case that technically sophisticated, highly complex finding based on complicated structural analyses were reported (Mundel, 1985: 254)--thereby, causing confusion among nontechnicians as to what was tested and what was found. Indeed, whether for better or worse, researchers using data from the income maintenance experiments generally appeared more concerned with communicating with their professional peers than with influencing public policy.

The communication of findings from the income maintenance experiments has, in fact, been explicitly criticized in a report by the General Accounting Office (1981), a report that suggests that lack of effective dissemination of these results may have been partially responsible for the failure of the Carter Administration's 1977 welfare reform proposal. In its report, GAO argued that had the findings been disseminated earlier and in a more organized fashion, and had they been summarized in layperson's terms, Congress might not have been so confused by them.[9]

There is little doubt that communication of findings from the work/welfare demonstrations was more effective than the dissemination of results from the income maintenance experiments. Some dissemination at the subnational level was virtually automatic since, in contrast to the income maintenance experiments, the work/welfare demonstrations were run by individual states. In addition, there was a single source of information about findings from the demonstrations: MDRC. MDRC's reports on each of these demonstrations are all written in a similar format. And the emphasis is on simple comparisons of differences between experimentals and controls. Moreover, each report contains an executive summary, which, although typically quite lengthy, is written in nontechnical language. Probably more important, several relatively brief, nontechnical summaries of overall findings from the demonstrations have also been disseminated (see Gueron,

---

[9]Neuberg (1988) has argued that the information communicated to policymakers and the general public concerning findings from the income maintenance experiments highly "distorted" what these experiments could actually say about the policies that were under consideration.

1986, 1987, and 1990). Baum (1989: 17-18) asserts that the 1987 summary, the appearance of which she refers to as a "scintillating omen," was especially important in the development of the 1988 Family Support Act: "Finding out whether the [work/welfare] demonstrations worked better than what we had before was important. Having the data in time to help shape and promote our legislative efforts was nothing short of amazing." Furthermore, MDRC, usually in the person of its president, Judith Gueron, frequently briefed interest and advocacy groups and Congressional staffers (Szanton, 1990: 13) and, in addition, testified on several occasions in Congressional hearings (Haskins, 1990 and Baum, 1989: 18).

Generalizability

Generalizability refers to whether research findings can be applied to different time periods or places. Thus, basic theoretical research is typically more generalizable than highly applied empirical research. In the context of social experimentation, an important related issue is whether the findings pertain to policies that differ somewhat from the ones directly tested. As should be apparent, the more generalizable a particular piece of research, everything else equal, the more likely it is to be used.

In principle, the income maintenance experiments are considerably more generalizable than the work/welfare demonstrations. The reason for this is that the former may be characterized as "response surface" experiments and the latter as "black box" experiments (see Orr, 1985). That is, the income maintenance experiments were designed in a way that, at least in principle, permitted estimation of the effects of variations in the basic parameters of the programs being tested--specifically, changes in the income transfer amount provided by a negative income tax program and variations in the program's benefit-reduction rate. Once obtained, estimates of these response surface parameters could be used to project the effects of any program that had the

basic features of a negative income tax, even if the transfer amounts and the program's benefit reduction rate differed substantially from those tested in the experiments.

The work/welfare demonstrations, in contrast to the income maintenance experiment, tested job search and work programs that could not be characterized by a small set of infinitely variable parameters. Thus, findings from these experiments were not well suited for extrapolation to programs that differed substantially from those tested. As Baum (1989: 22) put it, "Only certain strategies had been tested at a limited number of sites. Whether the same outcomes would occur in a national program with different requirements and a broader population of AFDC recipients was impossible to say."

The potential usefulness of social experiments that permit estimation of response surface parameters is perhaps best illustrated by the incorporation of parameters estimated from the Seattle-Denver Experiment into several different microsimulation models (see Betson and Greenberg, 1983; Haveman, 1987: Chapter 10). These models, which are based on microdata on large representative samples of the nation's households, have been used to predict the effects and costs of substantial changes in the nation's welfare system. They are particularly well suited for predicting who the winners and losers would be from various alternative policy variations. Microsimulation models that incorporated parameters from the Seattle-Denver Experiment were used extensively in designing and costing out the Carter administration's welfare reform proposals. Thus, micro-simulation allowed findings from the experiment to be extrapolated to the nation as a whole and to transfer programs other than those directly tested in the experiment. By permitting this, microsimulation became a bridge between the experiment and concrete-elaborative policy use. However, the importance of this bridge should not be exaggerated. Micro-simulation is complex, easily subjected to technical criticism that tends to undermine its credibility, and difficult to describe in readily accessible form to policymakers. Although it is fairly clear that

microsimulation and, consequently, the Seattle-Denver experimental findings had an impact on the design of the Carter proposal (Aaron and Todd, 1978), it is not evident that the simulation results had much, if any, impact on Congress's ultimate decision.

Relevance

Relevance concerns whether the needs of decisionmakers are met by the research (Leviton and Hughes, 1981: 527). If the research is perceived by decisionmakers as meeting their needs, they are obviously more likely to use it. The utilization literature suggests that research that focuses on manipulable variables, as social experiments do, tends to be viewed as more useful to decisionmakers (Beyer and Trice, 1982: 613). Moreover, the potential usefulness of social experimentation and other forms of evaluation is typically readily apparent. As Leviton and Hughes (1981: 537) have suggested, "evaluation differs from some other social research in that users are usually aware of the relevance of evaluations to policies and programs."

Relevance may interact with two of the utilization factors discussed earlier: generalizability and timeliness. For example, tradeoffs between relevance and generalizability often exist since research directed at a specific policy--for instance, a black box experiment that tests a particular program--may be highly relevant to assessing that policy, but less generalizable to other policies. Moreover, the less generalizable the research, the more perishable it may be (Beyer and Trice, 1982: 605). Thomas (1987: 52) appears to suggest this when she asserts that "Ironically, the more closely...research is geared to policy, the less use will be made of it. The reasons are not hard to find: the normal life of a research project is about three years with perhaps six months for writing up the results. Within that time the agenda for action will have changed."

The problem Thomas writes of is a real one. For example, none of the four income maintenance experiments tested a public service jobs component, but the Better Jobs and Income

Program included such a component. This, of course, reduced (but far from eliminated) the relevance of the income maintenance experiments to the Carter proposal. Nevertheless, Thomas appears to exaggerate the seriousness of the problem. Policy agendas are not necessarily completely revised every three and one-half years. For example, although the work/welfare demonstrations began in the early 1980s, results from these experiments appeared to be viewed as still highly relevant when Congress was considering the Family Support Act in the late 1980s. As previously mentioned, however, the time schedules of research and policy are relatively independent of one another. Thus, had the gap between initiating the work/welfare demonstrations (which, in fact, were highly program specific black box experiments) and the reappearance of welfare reform on the policy agenda been substantially longer than it actually was, results from the demonstrations might have been viewed as less relevant than they were.

Not surprisingly "relevance" is often defined in terms of a user's own narrow self-interest, rather than in terms of social interest. For example, Beyer and Trice (1982: 613), in their review of the utilization literature, point out that potential users have been found to view research more favorably if they perceive it as relevant to their own current job responsibilities. More seriously, but again not unexpectedly, it has also been found that in using research for symbolic purposes, users tend to find relevant only those results that conform to their prior conclusions (Beyer and Trice, 1982: 612-613). This suggests, for example, that findings that treatments under the income maintenance experiments had negative effects on labor supply and family stability were viewed as more relevant by persons arguing against the Nixon and Carter welfare reform efforts of the 1970s than by persons supporting these efforts. Similarly, findings that treatments under the work/welfare demonstrations tended to increase earnings and reduce welfare dependency, albeit modestly, were viewed as highly relevant by supporters of the Family Support Act of 1988.

However, although it is certainly true that both sets of experiments were used to score debating points--that is, for concrete-symbolic purposes--we suggested earlier that the work/welfare demonstrations resulted in concrete-persuasive use as well, while the income maintenance experiments did not. One possible explanation for this difference between the two sets of experiments revolves around the relevance of what they tested to political decisionmakers. For example, the central focus of the income maintenance experiments was the impact of income transfers on individual behavioral decisions affecting labor supply and family stability. But, as previously argued, in assessing the Family Assistance Plan and the Better Jobs and Income Program, behavioral effects were only one of several issues with which lawmakers were concerned, and, perhaps, not the most important. Thus, the focus of the income maintenance experiments, although not irrelevant, was somewhat tangential to the central interests of many policymakers. In contrast to the income maintenance experiments, the outcomes focused on in the work/welfare demonstrations seemed to address the concerns of lawmakers assessing the Family Support Act quite directly. For example, among the outcomes measured were the effects of the treatments on the probability of receiving welfare and on government budgetary expenditures, outcomes that were central to the interests of lawmakers.

CONCLUSIONS

We view this paper as an initial attempt to look at social experiments through the lens provided by the knowledge utilization literature. This initial glance suggests that there is indeed much to be gained by viewing social experiments through this lens. For example, the knowledge utilization literature provides explanations concerning when experiments are conducted, explanations that are plausible alternatives to the classical one that the decision to conduct a social experiment is made according to an essentially rational paradigm.

This literature also helps highlight the range of uses to which findings from social experiments can be put and the circumstances under which various types of uses are more or less likely. Recognition of the range of uses to which the results of social experiments are put is a possible antidote to the tendency of many proponents of social experiments to consider only concrete-substantive effects. Such a narrow view of the utilization of social experiments often puts such proponents in the position of proposing an experiment that they promise can be used instrumentally to determine decisively whether or not the tested program should be adopted. Since once the experiment is completed, such concrete-substantive effects are rarely apparent, this both reduces the credibility of social experiments and leaves their proponents demoralized. Recognizing the varied effects of social experiments, then, is likely to be beneficial for both the morale and credibility of proponents of social experiments.

Our examination of social experiments also suggests that Weiss's I-I-I (Ideology-Interest-Information) framework has some predictive value in determining when a social experiment is more or less likely to have effects. The greater the consistency among existing ideologies, information from other sources, and existing political interests, this model suggests, the less the likelihood of an experiment having policy impacts. At least with respect to concrete-persuasive effects, the contrast between the income maintenance experiments and the work/welfare demonstrations represents one piece of supporting evidence for this hypothesis. However, experience with the income maintenance experiments and the work/welfare demonstrations suggest that a high degree of I-I-I consistency does not appear to greatly reduce the likelihood of concrete-elaborative, concrete-symbolic, or the various forms of conceptual use. Applying this model to future proposed experiments appears to have value both in establishing realistic expectations and in designing the experiments.

One would anticipate on the basis of the I-I-I model, for example, that black box experiments are more amenable to concrete-substantive and concrete-persuasive use than response surface experiments. This is especially likely to be true when, as is frequently the case, a black box experiment tests an incremental policy change through existing government agencies. The very fact that the experimental treatment has been implemented by a regular government agency implies the absence of strong challenges to existing ideologies and political interests.

Response surface experiments tend to be highly complex in design and, hence, not easily administered by existing government agencies. They are probably better suited for conceptual use than black box experiments, however. As previously noted, findings from the latter are typically limited to a specific policy; they are not readily generalizable. Response surface experiments, on the other hand, may have at least some applicability to a wide variety of issues. Findings from the income maintenance experiment, for example, which are reported in most modern labor economics textbooks, appear to have considerably reduced uncertainty concerning the elasticity of the labor supply of men and women (Stafford, 1985). Thus, even if these experiments had little or no direct impact on the political fate of the tested NIT plans, they may have some future influence on a wide range of policies in which labor supply is an issue. This influence, however, is unlikely to be strong or direct; indeed, if it occurs, it may not even be recognized.

In the preceding section of the paper, we examined five factors, drawn from the research utilization literature, that are purported to influence the extent to which a specific set of research findings are used by policymakers: credibility; timeliness; communication and visibility; generalizability; and relevance. The income maintenance experiments and the work/welfare experiments were compared in terms of each of these factors. Since response surface experiments, by definition, are more easily generalized to different time periods and places than black box experiments, the income maintenance experiments appear superior to the work welfare

experiments in terms of this factor. It is not evident that the two sets of experiments rank very differently in terms of timeliness. Although the income maintenance experiments took considerably longer to complete than the work/welfare experiments, findings from the former were available in time for debate over the Program for Better Jobs and Income (but not at the time the Family Assistance Plan was proposed by the Nixon Administration) and findings from the latter were available at the time Congress considered the Family Support Act.

Ranked in terms of the remaining three factors--credibility, communication and visibility,and relevance--the work/welfare experiments appear to dominate the income maintenance experiments. In the paper, we examined a number of reasons for this. For example, the income maintenance experiments were analyzed by a diversity of researchers and findings were reported by a number of different sources. In contrast, only one institution was involved in analyzing and reporting findings from the work/welfare experiments--MDRC. In addition, the policies tested under the work/welfare experiments were considerably more incremental in nature than those tested under the income maintenance experiments. Moreover, as compared to the income maintenance experiments, findings from the work/welfare experiments were less scrutinized; more consistent with prior expectations; based on simpler estimation techniques and, hence, more readily understood by policymakers; more consistent across individual experiments; more effectively communicated to policymakers; and focused on outcomes that were more central to the primary interests of policymakers.

Thus, it would appear that there are a number of reasons, in addition to the constellation of ideologies, interests, and information from sources other than research, why findings from the work/welfare demonstrations had concrete-persuasive use while findings from the income maintenance experiments did not. However, several of the points made in the preceding paragraph suggest some potentially important issues for researchers that, while beyond the scope

of the present paper, should be raised. For example, it appears that research findings that are tests of incremental policy changes, subject to little scrutiny, reported by a single voice, and based on fairly simple estimation techniques may be more likely to be used in policymaking than research findings for which the opposite conditions hold. Furthermore, it seems that researchers can enhance the probability that their findings will be used by policymakers by devoting considerable resources to communicating with such persons, even if this comes at the expense of communicating with their professional peers. Consideration should be given to whether the long-run goals of either the policy process or science are best served by this state of affairs.

In concluding, we emphasize that this paper represents only an initial glance at social experiments through the lens provided by the knowledge utilization literature. We have looked at only two (though certainly two of the more prominent) sets of social experiments. Moreover, we have relied upon only a few primary sources and several secondary sources. Thus, this paper should be viewed as exploratory research. Our findings are best interpreted as representing hypotheses, rather than firm conclusions. To test these hypotheses, it is necessary to expand both the number of experiments upon which we focus and the number of original sources--perhaps, through interviews and questionnaires--from which we obtain data concerning the policy effects of various sets of social experiments.

References

Aaron, Henry J. Why Is Welfare So Hard to Reform? Washington, D.C.: Brookings Institution, 1973.

Aaron, Henry J. Politics and the Professors. Washington, D.C.: Brookings Institution, 1978.

Aaron, Henry J. "Six Welfare Questions Still Searching for Answers." The Brookings Review (Fall 1984): 12-17.

Aaron, Henry J. "Comment." In Jerry A. Hausman and David A. Wise, eds., Social Experimentation. Chicago: University of Chicago Press, 1985.

Aaron, Henry J. "Social Science Research and Policy." Journal of Human Resources 25 (Spring 1990): 276-280.

Aaron, Henry J. and John E. Todd. "The Use of Income Maintenance Experiment Findings in Public Policy, 1977-1978." Industrial Relations Research Association, Proceedings of the 31st Annual Meeting, Chicago, August 1978.

Baum, Rikki. Untitled, unpublished manuscript, Progressive Policy Institute, Washington, D.C., 1989.

Betson, David and David Greenberg. "Uses of Microsimulation in Applied Poverty Research." In Richard Goldstein and Stephen M. Sacks, eds., Applied Policy Research. Totowa, New Jersey: Rowman and Allanheld, 1983.

Beyer, Janice M. and Harison M. Trice. "The Utilization Process: A Conceptual Framework and Synthesis of Empirical Findings." Administration Science Quarterly 27 (December 1982): 591-622.

Bozeman, Barry. "The Credibility of Policy Analysis: Between Method and Use." Policy Studies Journal 14 (June 1986): 519-539.

Bryk, Anthony, ed. Stakeholder-based Evaluation. San Francisco: Jossey-Bass, 1983.

Burke, Vincent J. and Vee Burke. Nixon's Good Deed: Welfare Reform. New York: Columbia University Press, 1974.

Burtless, Gary. "The Effect of Reform on Employment Earnings and Income." In Phoebe H. Cottingham and David T. Ellwood, eds., Welfare Policy for the 1990s. Cambridge, Massachusetts: Harvard University Press, 1989.

Burtless, Gary. "The Economist's Lament: Public Assistance in America." Journal of Economic Perspectives 14 (Winter 1990): 57-78.

Burtless, Gary and David Greenberg. "Inferences Concerning Labor Supply Behavior Based on Limited Duration Experiments." American Economic Review 72 (June 1982): 489-497.

Cain, Glen G. and Douglas A. Wissoker. "A Reanalysis of Marital Stability in the Seattle-Denver Income Maintenance Experiment." American Journal of Sociology 95 (March 1990): 1235-1269.

Cogan, John F. Negative Income Taxation and Labor Supply: New Evidence from the New Jersey-Pennsylvania Experiment. Report R-2155-HEW. Santa Monica, California: The Rand Corporation, February 1978.

Coleman, James. "The Uses of Social Science in the Development of Public Policy." Unpublished manuscript, University of Chicago, 1979.

Coursey, David H. "Information, Credibility, and Decision Outcomes: An Experimental Examination of Credibility Logic." Ph.D. dissertation, Syracuse University, in progress.

Cox, Gary B. "Managerial Style: Implications for the Utilization of Program Evaluation Information." Evaluation Quarterly 1 (1977): 499-508.

Coyle, Dennis J. and Aaron Wildavsky. "Social Experimentation in the Face of Formidable Fables." In Alicia H. Munnell, ed., Lessons from the Income Maintenance Experiments. Boston: Federal Reserve Bank of Boston, 1986.

Elmore, Richard F. "A Political Scientist's View of the Income Maintenance Experiments." In Alicia H. Munnell, ed., Lessons from the Income Maintenance Experiments. Boston: Federal Reserve Bank of Boston, 1986.

Feldman, Martha S. "Producing Policy Papers." In Advances in Information Processing in Organizations. Volume 2. Greenwich, Connecticut: JAI Press, 1985.

Feldman, Martha. Order without Design: Information Production and Policy Making. Stanford, California: Stanford University Press, 1989.

Greenberg, David, Robert Moffitt, and J. Friedmann. "Effects of Underreporting on the Estimation of Work Effort in the Gary Income Maintenance Experiment." Review of Economics and Statistics 63 (November 1981): 581-589.

Greenberg, David H. and Philip K. Robins. "The Changing Role of Social Experiments in Policy Analysis." Journal of Policy Analysis and Management 5 (1986): 340-362.

Greenberg, David and Michael Wiseman. What Did the OBRA Demonstrations Do? Conference Paper, Madison, Wisconsin, Institute for Research on Poverty, 1990.

Gueron, Judith M. Work Initiatives for Welfare Recipients. New York: Manpower Demonstration Research Corporation, March 1986.

Gueron, Judith M. "Reforming Welfare with Work." Occasional Paper No. Two, Ford Foundation Project on Social Welfare and the American Future. New York: The Ford Foundation, 1987.

Gueron, Judith M. "Work and Welfare: Lessons from Employment Programs." Journal of Economic Perspectives 4 (Winter 1990): 79-98.

Hanushek, Eric A. and David L. Weimer. "Deciding What to Do: Analytical Strategies and the Policy Environment." Unpublished manuscript, Rochester University, Department of Economics, May 1989.

Haskins, Ron. "Congress Writes a Law: Research and Welfare Reform." Unpublished manuscript, Washington, D.C., 1990.

Haveman, Robert H. Poverty Policy and Poverty Research: The Great Society and the Social Sciences. Madison, Wisconsin: University of Wisconsin Press, 1987.

Heclo, Hugh. "Issue Networks and the Executive Establishment." In Anthony King, ed., The New American Political System. Washington, D.C.: American Enterprise Institute, 1978.

Jenkins-Smith, Hank C. Democratic Politics and Policy Analysis. Pacific Grove, Calif.: Brooks/Cole Publishing Company, 1990.

Kingdon, John W. Agendas, Alternatives, and Public Policies. Boston: Little Brown and Company, 1984.

Leviton, Laura C. and Edward F.X. Hughes. "Research on the Utilization of Evaluations: A Review and Synthesis." Evaluation Review 5 (August 1981): 525-548.

Lynn, Lawrence and David Whitman. The President as Policymaker: Jimmy Carter and Welfare Reform. Philadelphia: Temple University Press, 1981.

Majone, Giandomenico. Evidence, Argument and Persuasion in the Policy Process. New Haven, Connecticut: Yale University Press, 1989.

Mandell, Marvin B. "User-Involvement in the Design and Selection of Evaluation Studies: A User-oriented Framework and Its Assessment." Knowledge 5 (June 1984): 419-445.

Mandell, Marvin B. "The Consequences of Improving Dissemination in Garbage-Can Decision Processes: Insights from a Simulation Model." Knowledge 9 (March 1988): 343-361.

Mandell, Marvin B. "A Simulation-Based Assessment of the Value of Enhancing the Credibility of Policy Analysis." Knowledge in Society: The International Journal of Knowledge Transfer 2 (Summer 1989): 39-56.

Mandell, Marvin B. and Vicki L. Sauter. "Approaches to the Study of Information Utilization in Public Agencies: Problems and Pitfalls." Knowledge 6 (December 1984): 145-164.

March, James G. and Johan P. Olsen. Ambiguity and Choice in Organizations. Bergen, Norway: Universitetsforlaget, 1976.

Mintzberg, Henry. "Managerial Work: Analysis from Observation." Management Science 18 (1971): B97-B110.

Moffitt, Robert and Kenneth Kehrer. "The Effect of Tax and Transfer Programs on Labor Supply: The Evidence from the Income Maintenance Experiments." In Ronald Ehrenberg, ed., Research in Labor Economics, Volume 4. Greenwich, Connecticut: JAI Press, 1981.

Moynihan, Daniel P. The Politics of a Guaranteed Income: The Nixon Administration and the Family Assistance Plan. New York: Random House, 1973.

Mundel, David S. "The Use of Information in the Policy Process: Are Social-Policy Experiments Worthwhile?" In Jerry A. Hausman and David A. Wise, eds., Social Experimentation. Chicago: University of Chicago Press, 1985.

Munnell, Alicia H., ed. Lessons from the Income Maintenance Experiments. Boston: Federal Reserve Bank of Boston, 1986.

Neuberg, Leland G. "Distorted Transmission: A Case Study in the Diffusion of Social 'Scientific' Research." Theory and Society 17 (1988): 487-525.

Orr, Larry L. "Discussion." In Jerry A. Hausman and David A. Wise, eds., Social Experimentation. Chicago: University of Chicago Press, 1985.

Palmer, John L. and Joseph A. Pechman, eds. Welfare in Rural Areas: The North Carolina-Iowa Income Maintenance Experiment. Washington, D.C.: The Brookings Institution, 1978.

Patton, Michael Q., Patricia S. Grimes, Kathryn W. Guthrie, Nancy J. Brennan, Barbara D. French, and Dale A. Blyton. "In Search of Impact: An Analysis of the Utilization of Federal Health Evaluation Research." In Carol H. Weiss, ed., Using Social Research in Public Policy Making. Lexington, Massachusetts: Lexington Books, 1977.

Pechman, Joseph A. and Michael P. Timpane, eds. Work Incentives and Income Guarantees: The New Jersey Negative Income Tax Experiment. Washington, D.C.: Brookings Institution, 1975.

Reischauer, Robert D. "Discussion." In Alicia H. Munnell, ed., Lessons from the Income Maintenance Experiments. Boston: Federal Reserve Bank of Boston, 1986.

Rossi, Peter and Katharine C. Lyall. Reforming Public Welfare: A Critique of the Negative Income Tax Experiment. New York: Russell Sage Foundation, 1976.

Sabatier, Paul. "The Acquisition and Utilization of Technical Information by Administrative Agencies." Administrative Science Quarterly 23 (September 1978): 396-417.

Sproull, Lee and Patrick Larkey. "Managerial Behavior and Evaluator Effectiveness." In Herbert C. Schulberg and Jeanette M. Jerrell, eds., The Evaluator and Management. Beverly Hills, California: Sage, 1979.

Stafford, Frank P. "Income-Maintenance Policy and Work Effort: Learning from Experiments and Labor-Market Studies." In Jerry A. Hausman and David A. Wise, eds., Social Experimentation. Chicago: University of Chicago Press, 1985.

Szanton, Peter. "The Remarkable Quango: Knowledge, Politics and Welfare Reform." Unpublished manuscript, New York, New York, 1990.

Thomas, Patricia. "The Uses of Social Research: Myths and Models." In Martin Bulmer, ed., Social Science Research and Government: Comparative Essays on Britain and the United States. New York: Cambridge University Press, 1987.

Thompson, Mark S. Decision Analysis for Program Evaluation. Cambridge, Massachusetts: Ballinger Publishing Company, 1982.

U.S. General Accounting Office. Income Maintenance Experiments: Need to Summarize Results and Communicate the Lessons Learned. Report Number HRD 81-46. Washington, D.C.: GPO, April 17, 1981.

Webber, David J. "Explaining Policymakers' Use of Policy Information: The Relative Importance of the Two-Community Theory versus Decision-Maker Orientation." Knowledge 7 (March 1986): 249-290.

Weiss, Carol H. "The Many Meanings of Research Utilization." Public Administration Review 39 (September-October, 1979): 426-431.

Weiss, Carol H. "Knowledge Creep and Decision Accretion." Knowledge: Creation, Diffusion, Utilization 1 (March 1980): 381-404.

Weiss, Carol H. "Ideology, Interests, and Information: The Basis of Policy Positions." In D. Callahan and B. Jennings, eds., Ethics, Social Science, and Policy Analysis. New York: Plenum, 1983.

Weiss, Carol H. "Increasing the Likelihood of Influencing Decisions." In L. Rutman, ed., Evaluation Research Methods: A Basic Guide. 2nd edition. Beverly Hills, California: Sage, 1984.

Weiss, Carol H. "Congressional Committee Staffs (Do, Do Not) Use Analysis." In Martin Bulmer, ed., Social Science Research and Government: Comparative Essays on Britain and the United States. New York: Cambridge University Press, 1987.

Weiss, Carol H. and Michael J. Bucuvalas. "Truth and Utility Tests: Decision Makers' Frame of Reference for Social Science." American Sociological Review 45 (April 1980): 302-313.

Whiteman, David. "Reaffirming the Importance of Strategic Use: A Two-dimensional Perspective on Policy Analysis in Congress." Knowledge 6 (March 1985): 203-224.