

We review the arguments for and against randomized field experiments design to address important questions of social policy. Based on this review, we make a number of recommendations about how the use of randomized field experiments might be fostered.

SOCIAL POLICY EXPERIMENTATION

A Position Paper

RICHARD A. BERK

University of California, Santa Barbara

ROBERT F. BORUCH

Northwestern University

DAVID L. CHAMBERS

University of Michigan

PETER H. ROSSI

University of Massachusetts

ANN D. WITTE

Wellesley College

I: INTRODUCTION

In Donald T. Campbell's justly famous article, "Reforms as Experiments" (1969), an explicit link was proposed between rigorous experimental methods and social reforms:

The United States and other modern nations should be ready for an experimental approach to social reform, an approach in which we try out new programs designed

AUTHORS' NOTE: Funding for this position paper was provided by the John D. and Catherine T. MacArthur Foundation. We are also indebted to Lawrence J. Hubert, Otis

EVALUATION REVIEW, Vol. 9 No. 4, August 1985 387-429

© 1985 Sage Publications, Inc.

to cure specific social problems, in which we learn whether or not these programs are effective, and in which we retain, imitate, modify, or discard them on the basis of apparent effectiveness [Campbell, 1969: 409].

Experimental methods had, of course, long been popular in those natural and social sciences with laboratory traditions. And the technical advances made by Gosset, Fisher, Yates, and others had permitted rigorous experimentation to be undertaken outside of laboratory settings. Likewise, the notion of social reform was hardly new. There had even been a number of earlier attempts to apply experimental methods to estimate the effect of changes in public policy.

Campbell's contribution was the argument that social reforms and experimental methods should be *routinely* adjoined. A commitment to an experimental *philosophy* would make it possible to be innovative and cautious at the same time; it would be possible to try new things in overtly tentative manner. A commitment to the experimental *method* would allow *factual* questions about the impact of the social reforms to be addressed within the best scientific traditions. Thus, whether interventions such as Head Start, peak-load pricing, Job Corps, or CETA "worked" was an empirical question to be explored with the most demanding of scientific procedures. The same logic applied in principle to any change in policy: Revisions of a penal code, deregulation of particular industries, alterations in school curricula, the implementation of environmental legislation, and many others.

Since Campbell's statement 15 years ago, the experimental approach to social reform has had an uneven history. A number of major experiments were launched, and much was learned. For instances, we now know that residential consumers of water and electricity can be induced to "conserve" when the price per unit is increased to more closely approximate the true marginal cost. We also know that prison rehabilitation programs operating in the 1960s did not on the average demonstrably reform criminals.

However, meaningful experiments also proved harder to implement than many had expected: one could not simply move laboratory

Dudley Duncan, and Sarah Fenstermaker Berk for their comments on an earlier draft of this report. Finally, some of the better ideas in the article were formulated in general discussions the senior author had with members of the Social Science Research Council's Advisory Committee on Social Indicators.

technique into the field. For example, it was often difficult in practice to maintain tight control over implementation of a research design. In one large study (Rossi and Lyall, 1976), for instance, the experimental treatment involved providing income subsidies to randomly selected households, at the same time providing no support for another set of randomly selected, comparison households. These comparison households were also not eligible for other kinds of public assistance. Unfortunately, midway through the experiment, local officials changed the eligibility rules for AFDC payments so that many of the comparison households now qualified for AFDC support. Consequently, the experimental subsidies no longer effectively distinguished the experimental households from many of these comparison households; both were often receiving approximately the same level of income.

The use of experimental results by policymakers has also been uneven. Some experiments have had a clear and direct impact on important decisions. For example, a recent experiment on police responses to domestic violence incidents has already affected law enforcement practices in a number of major departments (Sherman and Berk, 1984). More commonly, experimental results have been ignored or sometimes even misused. No doubt one major obstacle has been poor communication between the scientific community and public officials. Another obstacle has been a lack of incentives for both researchers and policymakers to invest heavily in experimental approaches to social reform.

In our view, these and other difficulties can be attributed significantly to the growing pains of a new endeavor. Researchers have gradually learned to promise less and deliver more. Policymakers are gradually learning what experiments can and cannot accomplish. The more general point is that reforms *can* be introduced as experiments, and we have learned a great deal about when experimental procedures make sense, how laboratory methods should be altered when moved into the field, and experimental results can be made more responsive to the needs of policymakers. However, we are still a very long way from an effective marriage between experimental methods and assessments of public policy.

In this position paper, therefore, we argue for a concerted effort to facilitate social policy experimentation. By social policy experiments, we have in mind the application of rigorous research designs to determine the impact of changes in social policy.¹ Thus, a social policy

experiment is *an effort to introduce social change in a way that allows one to effectively discern the net effect of the change on important social outcomes*. We require, as a result, that the "treatment" be a legitimate and manipulable instrument of public policy. Changes in our tax laws would qualify whereas altering the age mix of our population would not (because it is not manipulable in practice). We also require that resources be provided "up front" for estimating the impact of the policy change. One cannot have a social policy experiment without the resources to conduct the research. Finally, we require that existing theory and knowledge be exploited; we leave to others undirected "tinkering" with the system. Thus, we would favor in principle an experimental treatment of transfer payments designed from micro-economic theory, but within the context of social policy experiments disapprove of transfer payments justified solely as a means of making poor people less poor.

In summary, we believe that the time is right to launch a major *and* coordinated effort in social policy experimentation. We try to make the case in this position paper that social policy experiments can produce good science and good social policy; it is possible to do good and do it well.

The rest of our position paper is organized into four sections: section II provides some general background on the nature of social policy experiments; section III addresses social policy experiments in more depth through three experiments on a single topic; and section IV provides a definition of social policy experiments and elaborates on that definition. Finally, section V describes a number of concrete proposals that may further social policy experimentation.

II: SOME BACKGROUND ON SOCIAL POLICY EXPERIMENTS

The term "social policy experiment" can connote a variety of activities from seat-of-the-pants innovations in public policy to scientific experiments on social policy tested in the field. Although we will shortly be far more specific, it will be apparent that we are concerned with the use of experimental procedures to determine the impact of social interventions having significant implications for public policy. We begin with four examples.

EXAMPLES OF SOCIAL POLICY EXPERIMENTS

In the early 1960s, there was a growing concern that proposals to provide even modest income subsidies to the poor would reduce motivation to find and keep jobs. One charge was that "handouts" undermined the "work ethic." In response, several experiments were launched to determine if members of households that were given monthly cash grants at various levels (including no payments) varied as a consequence in their labor force participation. A second intervention involved the rate at which these payments were reduced in response to earnings from employment. In order to ensure that households in the different groups were as comparable as possible, households were assigned at random to the experimental (e.g., payment) and control (e.g., nonpayment) conditions. In fact, the work disincentives produced by the payments were on the average rather modest, and most of the labor force reductions came from households' "secondary workers" such as teenagers or women with small children (Rossi and Lyall, 1976). That is, work effort was reduced especially for those household members who presumably had other pressing demands on their time (i.e., school and child care, respectively). One implication was that a "negative income tax" to assist the poor and near poor was arguably less problematic (in terms of work disincentives) than many had predicted.²

At about the same time, prison rehabilitation programs were coming under attack as ineffective or worse. Criticisms from the right focused on administrators and programs that "coddled" prisoners. Criticisms from the left centered on civil liberties violations that rehabilitation programs necessarily engendered (e.g., the lack of due process in decisions about when "rehabilitation" was complete). As a step toward moving beyond impressionistic evidence, Kassbaum et al. (1971) compared the postprison behavior of convicts randomly assigned to group therapy with the postprison behavior of convicts randomly assigned to the usual prison programs. Arrests for new crimes were of particular concern. Parole records revealed no significant difference between the two groups, suggesting that nothing was gained from exposure to the group therapy, at least as delivered in this study. Part of the problem was the difficulty of delivering meaningful therapy using prison personnel operating behind bars. In any case, the study allowed both the Right and the Left to continue their assault on rehabilitation, fueled by the idea that "nothing works."

A bit more recently, there was interest in developing a Spanish version of *Sesame Street* for use in Mexico and other Latin American countries. It was not clear, however, whether the television program would have beneficial effects sufficient to justify its costs. It was also not clear whether all children would benefit about the same amount, or whether children who were already better equipped for school would benefit more than children who were less well equipped. Such differential responsiveness to the program might or might not improve school performance averaged over all children, but it would certainly increase the *gap* between the fast and slow learners. There was some debate about whether programs that fostered inequality between groups of school children were a good idea and deserving of support with taxpayer's money. To answer these and other questions, an experiment was launched in Mexico City to determine if children randomly assigned to watch over 100 episodes of *Plaza Sesamo* scored higher on a variety of achievement tests than children randomly assigned to watch an equivalent number of cartoon shows and other noneducational programs. On the average, the children exposed to *Plaza Sesamo* did better than the children exposed to other kinds of programs, especially on those tests most closely related to the stated goals of *Plaza Sesamo*. Apparently, all children benefited on the average, but as more talented children gained greater amounts than less talented children, the gap between fast and slow learners was increased (Diaz-Guerrero et al., 1976).

In the middle 1970s, utility companies were becoming increasingly worried about the "peak load" problem. They found that the demand for electricity varied enormously over the course of the day and over the course of the year. Providing for high-demand periods meant building new power plants costing hundreds of millions of dollars. The alternative was potential blackouts during periods of especially high demand. But even if such capacity were built, there would be many slack periods when the new capacity would go unused. Consequently, an experiment was initiated to determine if by raising the residential price of electricity during the peak load hours, consumers would shift substantial amounts of their demand for electricity to other times of the day. Households were assigned at random to different price structures, and on the average, consumption dropped by about 5% for a 10% increase in price, (Aigner and Hausman, 1980). In other words, it seemed possible to overcome at least part of the peak load problem by altering the price

structure. One implication was that perhaps pricing strategies should be tried before commitments for new power plants were made.

The four studies just described represent "social policy experiments" as we will use the term. Our primary aim in the pages ahead is to argue the case for social policy experiments and suggest ways in which foundations, government agencies, and even private enterprise might make a larger commitment to scientifically sound social experiments designed to inform public policy. Thus, we have written something far closer to a position paper than a literature review. Readers interested in the latter can find excellent treatments in any of a number of accessible sources (e.g., Reiken and Boruch, 1974; Cook and Campbell, 1979; Rossi and Freeman, 1983; Ferber and Hirsch, 1983; Tanur, 1983; Fienberg et al., 1984).

A FIRST PASS OVER SOME IMPORTANT ISSUES

People experiment all the time. Common sense quite properly counsels that in order to determine the consequences of some action, trying it out can be a pretty good idea. Implicit are three important notions: (1) speculations about how the world works can be meaningfully informed by reference to observable phenomena, (2) one must intervene in "business as usual" in order to get a clear fix on cause and effect, and (3) any measures of effect are inherently comparative. Whether anything is different after an intervention requires a benchmark of how things would have been in the absence of the intervention. Thus, we test-drive cars before making a purchase, grant tenure at universities only after a lengthy probationary period, and date before deciding whether or not to marry.

Over the past two centuries, common-sense views of experimentation have been substantially clarified and expanded. We now know, for example, that inferences about the relationship between cause and effect must take into account the full range of possible explanations for why a particular outcome materialized. Whether "New Math," for instance, really improves mathematical reasoning requires that the impact of New Math be separated from the impact of normal maturation in students' mathematical aptitudes. Likewise, whether psychotherapy really works requires that the impact of the therapy be distinguished from a remission of symptoms that would have occurred without intervention. Or,

whether the current economic recovery can be attributed to supply side economics requires that the independent impetus of changes in consumer spending (among other things) first be factored out.

There is now a widespread consensus among statisticians that for questions of causal inference some research strategies are more illuminating than others. In particular, experiments in which subjects are assigned at random to various treatments are the procedure of choice when causal inference is the dominating concern.³ Random assignment ensures that experimental and control groups are on the average equivalent before the treatments are introduced. Hence, fair comparisons between the outcomes of the experimental and control conditions can be made. Moreover, under random assignment in "factorial" designs, it is possible to test the impact of several different kinds of treatments. One might explore, for instance, whether conservation appeals and price increases each made independent contributions to reductions in the use of electricity during peak hours and whether one was more effective when the other was already in place. That is, the price effects might be enhanced once customers were sensitized by appeals to conserve.

In similar fashion, it is now widely recognized that generalizations from a particular experiment to other settings, subjects, and times are often problematic. Thus, New Math might work for students in small classes, but not for students in large classes. Similarly psychotherapy might be effective for individuals who volunteer for treatment, but not for individuals who are coerced into treatment by court order. Supply side economics might have worked in the 1920s, but not now. And just as with current views on causal inference, there is a virtual consensus that some research strategies are stronger than others. In particular, sound generalizations follow more readily from samples drawn at random from known populations coupled with packages of experiments in which one can determine the degree to which initial findings are routinely replicated.

Finally, it is now generally accepted that all experiments are subject to chance forces. That is, an apparent tendency for one or more interventions to work better than others may reflect nothing more than the "luck of the draw." For example, student performance on standardized mathematics tests can fluctuate from day to day in an unpredictable fashion due to a host of small effects unrelated to what is taught in the classroom (e.g., the amount of sleep the student had the night before). Consequently, it is possible that superior performance by students

learning New Math may be a fluke. Likewise, psychopathology and economic indicators are well known to respond to random perturbations that may be confused with genuine responses to experimental interventions.

The role of chance substantially complicates inferences from experiments because one must try to account for the possibility that the effects one interprets as "real" are actually happenstance. And it is here that notions such as "statistical significance" take on special import; one must formally be able to assess the probability that a given finding (or set of findings) is a chance outcome. Note that there are two kinds of chance errors: One may falsely find that an intervention does not work or one may falsely find that an intervention does work.

These and other considerations lead naturally to a definition of social policy experiments that will suffice until later, when we shall be more complete. By a social policy experiment we shall initially mean a *planned effort to produce unbiased, reasonably precise estimates of the effect of a social intervention or change. The plan involves identification or construction of a control condition that will yield, compared to other procedures, substantially less equivocal, quantifiable statements about the effect.* This definition excludes many kinds of applied social research (more on that later, as well) such as political polling, census enumerations in different political jurisdictions, and monthly surveys on labor force participation. None of these addresses the effects of a particular social intervention or includes a control condition to provide a baseline.

Our point more generally is that common-sense experimental methods have been practiced throughout recorded history, and in the recent past, experimental methods have been dramatically improved. Among those who believe that scientific methods can make a useful contribution to policy formation and testing, there is little debate about the overall merits of these procedures. Yet, modern experimental methods are used very selectively. They are of course, important tools for basic research, especially in the natural sciences. They are also used routinely by social scientists working in laboratory settings. In addition, modern experimental methods are from time to time used in applied research, especially in biomedical research. Recall, for example, the massive trials of the Salk vaccine 30 years ago and the recent and highly publicized experiment on cholesterol and heart disease. However, alternative social policies are rarely subjected to rigorous experimental scrutiny. For example, there are no rigorous experimental studies on the impact of the minimum wage, the deterrent value of longer prison terms,

how microcomputers affect worker productivity, whether the price mechanism might be widely used to discourage water and air pollution, whether entitlement programs really engender "welfare dependency," how private education compares to public education, how lowering the speed limit affects automobile fatalities, and many others. Rigorous experiments in social policy are the exception, not the rule.

There are several probable reasons why experimental methods have not become more central to the policy-making process. First, there are few incentives for public officials to undertake experiments today. Launching an experiment is an admission of ignorance, and there is little political mileage to be made in the search for answers. We expect our leaders to have the answers. Moreover, the political process too often settles questions of fact by ideological fiat. Can an old-fashioned liberal dare to find out whether unions really improve the worker's lot? Can an old-fashioned conservative dare to find out whether deregulation really leads to greater efficiency? Finally, experiments take time, a very scarce commodity in the political process. The recent cholesterol study, for instance, took 10 years to complete. Ten years can cross four presidential administrations.

For these and other reasons, rigorous experiments in social policy are rarely embraced (let alone funded) by public officials. This does not prevent these officials from routinely making statements of fact as if they were based on sound scientific knowledge. This also does not mean that social experiments are necessarily incompatible with the political process; indeed, among our goals is to suggest ways in which a better relationship might be forged between politicians and social experimenters.

Second, although much of the necessary technical expertise can be found in universities and other research institutions, there are few academic incentives to engage in social policy experiments (or applied research more generally). As we have argued elsewhere at some length (Berk, 1981), applied research is typically viewed as a second-class undertaking and rewards such as tenure are reserved primarily for those who do not soil their hands with practical concerns. For example, it is almost impossible to publish applied sociological research in any of the major sociological journals. But, these and other obstacles need not prevent academics from becoming more involved in social policy experiments; indeed, later we will suggest some ways in which this might be accomplished.

Third, the social policy experiments undertaken over the past two decades have been of uneven quality, relevance and cost-effectiveness

(Ferber and Hirsch, 1983). The record of research actually conducted does not provide a basis for a ringing endorsement of all social experiments. Part of the problem has been unrealistic expectations about the ease with which sound experiments could be undertaken. We now know that the implementation of proper experimental procedures is at least as important as the design of experiments. Both the randomization and the treatment itself must be successfully put in place.

Another part of the problem has been naiveté about how compelling the results of any empirical research could be to political decision makers. We now understand that there cannot be a single definitive experiment in social policy studies. All beliefs about what works are held tentatively and subject to many caveats. The findings from social policy experiments are certainly no different in principle. The advantage of rigorous experiments is that the caveats are usually fewer and far less damning. Hence, the proper benchmark for social experiments is not unassailable knowledge, but the information that would have *otherwise been available*. Alternatively put, is the policy process now substantially *better* informed with the experimental results in hand?

And yet another part of the problem has been that many of the earlier social policy experiments addressed questions that were in retrospect not easily answered within an experimental context. We are more humble now; we have a better sense about *when* to experiment. In short, one might conclude from the experience of the past two decades that social experiments are too often of questionable value. We believe that this overlooks the many successes of social experimentation and fails to take account of the enterprise's growing pains.⁴

Fourth, the nature of social policy experiments is often misunderstood. There are claims, for instance, that rigorous experimentation is typically unethical. Of course, serious ethical problems can and do arise, yet many experiments of great utility can be conducted that will meet strict ethical standards (Bennet and Lumsdaine, 1975; Boruch et al., 1978; Boruch et al., 1979; Boruch and Shadish, 1982). There are also claims that on grounds of cost-effectiveness, rigorous experimentation is too dear. However, such conclusions are not based on a careful assessment of the record (e.g., Spencer, 1980; Mosteller and Weinstein, 1984). For example, a *nonexperimental* study on the effectiveness of New York State's gun control laws is currently budgeted by the National Institute of Justice (NIJ) for over \$400,000. Yet, another NIJ project involving a *true experiment* on police practices in incidents of domestic violence has cost less than \$200,000, and instructive, rigorous experiments can often be undertaken for under \$50,000 (a

typical award by the National Science Foundation, Division of Biological and Social Sciences).

One must also not forget that even very expensive social policy experiments may be cost-effective when millions of tax dollars are at stake. For example, in 1980, the health-care budget in the United States was approximately \$200 billion (Mosteller and Weinstein, 1984). Consequently, experiments totaling \$1 billion per year might well be cost-effective if health-care costs were reduced by only .5%. This is hardly a demanding criterion. For example, experiments on increasing the competition among health-care providers (e.g., allowing nurse practitioners and paramedics to provide more services previously requiring a physician) might easily pay for themselves.⁵

It is also very expensive to keep people in prison. In California, for instance, a year of incarceration in the State's crowded and understaffed prisons costs about \$15,000 per prisoner. With approximately 30,000 inmates, the annual bill is around \$450 million. This implies that social experiments in California costing \$450,0000 yearly would probably pay for themselves if they led to savings of only .1% of the annual prison budget. Again, this is hardly an unreasonable target. For example, if it could be shown that prisoners who were released only 30 days before the end of their effective sentences presented no greater danger to the public than those who served their full sentences, hundreds of thousands of tax dollars could be saved.

Finally, consider social problems on a far smaller scale. In the Southwest, the average residential consumer pays about \$25 a month for water. For a town of 100,000, therefore, the annual residential water bill would amount to approximately \$2.5 million. Local experiments costing the community \$125,000 a year would pay for themselves, if they reduced each residential consumer's budget by only 5%. For example, if an experiment costing under \$125,000 could show that local ordinances prohibiting watering of lawns in the heat of the day were typically followed without expensive enforcement mechanisms, the benefit-cost ratio might be very favorable.

In summary, the idea of experimenting in a generic sense seems to be part of human nature and formal experimental methods are now well developed. There is also ample evidence that rigorous experiments in social policy are often feasible, ethical, and cost-effective. Yet, social policy experiments are rarely initiated in order to determine what works. We believe that as a result, a very important policymaking tool is being underutilized. In the pages ahead, we will consider more deeply

the nature of social policy experiments, how they might enhance the policymaking process, and what public agencies, private foundations, and private enterprise might do to improve matters.

III: A CASE STUDY OF THREE SOCIAL POLICY EXPERIMENTS— POSTRELEASE BENEFITS FOR PRISONERS

The previous two sections of this article have considered social policy experiments in rather general terms. As we turn to a more detailed treatment, it will perhaps be instructive to begin with a more lengthy treatment of some examples. Consequently, we proceed to a description of three social policy experiments related to one another in the problems addressed, the interventions tried, and the outcomes explored.

STAGE 1: THE LIFE EXPERIMENT

The Manpower Development and Training act was passed in 1962. Amendments to the act in subsequent years added provisions for vocational training and other job-related services for individuals incarcerated in state and federal prisons. It was hoped that these programs would increase the chances of employment after release from prison. Evaluation of these programs, however, could find no impact on either employment or recidivism (e.g., Taggart, 1972).

Meanwhile, the Manpower Administration had commissioned a study of the postrelease experiences of federal prisoners. There were no surprises: ex-prisoners were typically released with very little "gate money" (often less than \$50) and had great difficulty finding jobs. One conclusion was that economic hardship contributed to the commission of new crime.

The failure of prison vocational training programs, the high unemployment rates of ex-prisoners, and the meager financial resources available at release led the Manpower Administration (later renamed the Employment and Training Administration) to consider new programs built around the idea of providing ex-prisoners with some sort of financial assistance to aid in the transition to life on the outside. It was hoped that such assistance would serve as a cushion, allowing ex-

offenders to search more effectively for a steady job and in the meantime, directly reduce the economic motivation for crime. The perspective was also explicitly experimental; a rigorous evaluation was contemplated.

After a very small and successful pilot study to determine the feasibility of a particular social policy experiment, the LIFE (Living Insurance for Ex-Offenders) experiment was designed. An initial step was the development of a rationale and a research design with which to deliver financial aid to ex-prisoners. In brief, because eligibility for unemployment benefits was (and is) determined from employment in the four to six quarters prior to an application for benefits, the vast majority of ex-prisoners necessarily were losing whatever eligibility they had. Hence, prisoners sentenced for felonies, who typically serve a year or more behind bars, could not possibly establish eligibility—irrespective of their employment prior to incarceration.

In response, one of the experimental conditions was to be an approximation of unemployment benefits. Individuals who could not find jobs would be eligible for \$60 a week for 13 weeks. A second experimental condition was job-placement services, much like those routinely provided by the U.S. Employment Service. A third experimental condition was a combination of the two “main” conditions. The fourth experimental (or control) condition was neither the benefits eligibility nor job counseling. Finally, approximately 400 prisoners were to be *randomly* assigned to each of the four conditions, and then followed for two years through interviews and official records (e.g., arrest reports) to measure any effects on postrelease behavior.

With the design in hand, the next step was to find a site to implement the experiment. For reasons that need not concern us here, the Maryland Department of Corrections agreed to cooperate, and the experiment was to be fielded in Baltimore. More important than the site were the criteria used to select prisoners to participate in the study. “High-risk” subjects were selected primarily because of a need to avoid floor effects; one could not determine which treatment most reduced the number of new crimes if very few crimes were committed overall. High risk was operationalized as males under 45 years of age with multiple convictions and at least one arrest for property crimes. No one addicted to alcohol or heroin was included, in part because the problems of addiction were assumed to be somewhat different from those of the “garden-variety” felon. There were other, less important, restrictions on participation that are described in Lenihan (1977) and Rossi et al. (1980).

Each of several analyses of the LIFE data revealed that financial aid reduced arrests on charges for property crimes. Approximately 30% of the individuals who did not receive benefits committed property crimes, compared to about 22% of those who received the benefits. The 8% difference was not a chance outcome by conventional criteria. On the other hand, no treatment effects were found for other kinds of crimes.

It was also believed by some that individuals receiving payments would be less likely to find and hold jobs. However, no such effects surfaced, in part because of a generous tax rate on earnings; it was possible to earn as much as \$150 a week and still receive some portion of the benefits.

Finally, when all of the findings were subjected to a benefit-cost analysis, it was estimated conservatively that were LIFE's findings to hold up in a regular program, 4 dollars would be saved for every dollar invested in the unemployment benefits. Clearly, the program would be cost-effective.

The LIFE results were certainly encouraging, but required at least three caveats. First, although the experimental effects were statistically significant by conventional standards, they were just barely so; chance effects were just barely being ruled out. Second, the experiment was administered by an energetic and dedicated research team. It was unlikely that should LIFE-like payments be administered by the personnel of state unemployment agencies, that the same level of effort would be forthcoming. Hence, it was possible that beneficial effects would fail to appear. Finally, the high-risk sample of ex-offenders was certainly not representative of all persons who might be eligible for the payments. Consequently, it was difficult to determine what might happen were unemployment benefits eligibility routinely made available to ex-prisoners. Clearly there was a need for more research.

STAGE 2: THE TARP EXPERIMENT

More research was also the recommendation of two gatherings of social scientists, government officials, and criminal justice professionals. This led to a larger and more realistic experiment dubbed TARP (Transitional Aid for Released Prisoners; see Rossi et al., 1980). Again there was to be random assignment to payment and nonpayment groups. Again there was to be a two-year follow-up. However, the unemployment benefits were to be administered by personnel from state unemployment offices, not specially recruited research staff. In addition, there was to be on the average a substantially larger tax on

earnings so that the payments would be dramatically reduced by rather modest earnings from employment. Finally, there were to be two experiments, one in Texas and one in Georgia, with about 1000 ex-prisoners in each, and virtually no restrictions on which ex-prisoners participated.

In contrast to the findings from the LIFE experiment, the TARP payments did *not* reduce overall property arrests or arrests for person crimes. As described in great detail in Rossi et al. (1980), a structural equation analysis of the experiment suggested that potential reductions in crime were swamped by an *increase* in new crimes stemming from unemployment *created* by the payments. In other words, there was apparently a tendency among those who received the benefits to reduce their labor force participation, at least in the short run.⁶ These individuals, in turn, were more likely to get into trouble. And the crimes committed by these individuals obscured the otherwise beneficial consequences of the payments. Recall that under LIFE, generous tax rates on earnings virtually eliminated any work disincentives. Under TARP, the work disincentives were substantial, and this led to what the TARP researchers called a "counterbalancing model": the payments had two *opposing* effects. (For another perspective, see Zeisel, 1982.)

The TARP study suggested that unemployment benefits could reduce both property and person crimes only if the work disincentives produced by the benefits could be affectively eliminated. And it was not clear precisely how that could be done, as the tolerance for error might not be very large. To complicate matters further, the inferences about appropriate levels of payments and the rate at which earnings would be taxed rested on findings from structural equations (i.e., statistical adjustments) that could not fully capitalize on the assets of random assignment to treatment and control groups. Hence, any causal inferences about the counterbalancing effects of unemployment benefits were more easily challenged than the overall conclusions that *as administered*, the payments did not reduce crime.

The content and timing of the TARP findings (corresponding with the beginning of the Reagan Administration) effectively ended any chance of federal initiative on transfer payments and crime. The LIFE and TARP experiments certainly did not show that providing unemployment eligibility to ex-offenders increased crime; the payments clearly did not make matters worse. There was also evidence that some form of transfer payments might well reduce crime. However, a clear mandate for a specific kind of program was not to be found. At least as

important, the values and premises that lay beneath the LIFE and TARP studies for other reasons fell out of fashion in Washington circles. Put a bit too simply, the stick replaced the carrot and interest turned to deterrence through punishment and incapacitation through imprisonment.

STAGE 3: CALIFORNIA SENATE BILL 224

At about the time that the TARP study was being designed, Peter Behr, a California state senator (Republican, Beverly Hills), heard about the LIFE experiment. He had long been associated with a number of moderate to liberal legislative programs and saw in the LIFE experiment the opportunity to reduce crime and save taxpayers a lot of money. He had read that programs built on the principles underlying LIFE could save at least 4 dollars for every dollar invested. (At that time, it cost about \$12,000 a year to keep a person in California incarcerated in a state facility.)

Drawing heavily on the LIFE experiment, Senate Bill 224 was introduced. The basic idea was to allow prisoners to establish eligibility for unemployment benefits from jobs held while doing time. Individuals working in prison jobs would be assumed to be earning the minimum wage of \$2.30 an hour (although they were actually earning about 20 cents an hour), and individuals who “earned” at least \$1500 over a 12-month interval would be eligible for unemployment benefits upon release. The size of the payments would depend on the level of earnings, but in practice, payments were likely to be a bit smaller than the \$60 per week provided by LIFE. In addition, all of the usual rules for receiving unemployment benefits were supposed to apply. For example, to receive the payments, an ex-prisoner had to be actively seeking work.

Despite Behr’s claims that the legislation would reduce crime and save tax dollars, it was opposed by conservative interests (e.g., professional law enforcement organizations, the California Chamber of Commerce). And not surprisingly, it was supported by liberal groups (e.g., the American Friends Service Committee). Clearly, ideology was carrying the day.

Eventually the bill became law, but in part to mollify the opposition, the legislation had two unusual features: There was a “sunset” clause giving the program a five-year life, and the program was to be subjected to a rigorous evaluation.⁷ In other words, the program was clearly

experimental and based on the evaluation results, would be completely reassessed at the end of five years. If the findings were favorable, the legislature presumably would be moved to introduce a *new* bill resurrecting the program.

All of this was far easier said than done. First, the legislature did not appropriate any funds to pay for the required evaluation. The California Department of Corrections was supposed to finance the research. In fact, the kind of evaluation demanded by the language of Senate Bill 224 was far more expensive than the Department of Corrections could afford.⁸ Second, at the time there was no technical expertise within the Department of Corrections to ensure a credible evaluation. And third, the need for technical sophistication was made more important by legislative language that all but precluded random assignment to experimental and control groups. It is now widely recognized by individuals experienced in social experimentation that the weaker the research design, the greater the need for elaborate statistical models.

The U.S. Department of Labor's Employment and Training Administration saved the day by providing the necessary funding. They too were interested in testing a program based on the LIFE experiment. Experienced researchers were recruited from the University of California, solving the technical expertise problem. However, a randomized experiment could not be implemented. It became necessary, therefore, to settle for a quasi-experimental approach resting on a regression-discontinuity design. For present purposes, the major point is that although the regression-discontinuity design is quite powerful, it is far more easily challenged than "true" experiments resting on random assignment to experimental and control groups.

The evaluation was to be finished and the results released to the legislature at the end of the program's fourth year. Six months earlier, however, an effort was made to scuttle the program before the fifth year could begin. The real motivations were primarily ideological; there was a genuine hostility toward ex-prisoners and a deep-seated feeling that they did not deserve any help from the state. The explicit rationale, however, was economic. The program was expensive, and there was alleged to be no evidence of effectiveness. Fortunately, the state senator who was carrying the bill said publicly that he was prepared to withdraw the legislation if there was any evidence that the program was working as promised. And preliminary analyses (rather dramatically) introduced during committee hearings shortly thereafter were optimistic. The legislation was withdrawn.

Six months later, the final report was available. It showed reductions in crime much like those found in the LIFE experiment and consistent with the TARP counterbalancing model. A conservative benefit-cost analysis suggested that the program saved California taxpayers about \$2000 per participant. (For more details, see Rauma and Berk, 1982; Berk and Rauma, 1983.) Based on these findings, an assembly bill was introduced to reinstitute the program, and it easily cleared the liberal-dominated committee in which it was initially heard. It foundered in the conservative-dominated assembly Ways and Means Committee (which reviews all spending measures). The argument was that the program perhaps saved money in the long run, but there was a fiscal crisis in California that could not wait for the long run to arrive. In all fairness, the state was facing a large deficit, which is prohibited by the California constitution. Times have again changed, however, and with the projected surplus California is now enjoying, it is likely that the bill will be reintroduced. Passage, of course, is another question.

SOME LESSONS

There are a number of lessons that might be extracted from the three studies just described. For our purposes, five stand out. First, the LIFE and TARP were exemplary applications of rigorous scientific methods to assess the impact of a social program (Cronbach, 1983). The evaluation of Senate Bill 224 was nearly as strong. Yet, none of the efforts produced social policy recommendations that were truly definitive; the general conclusion was that transfer payments might well reduce crime, but only if the work disincentives were not strong. Hence, although the information provided was far superior to the information that had earlier informed public policy, an ironclad case one way or the other was surely not made for extending unemployment benefits to ex-prisoners. In short, even exemplary social experiments can do no more than reduce uncertainty about the impact of various interventions. The reduction in uncertainty may be dramatic, but some uncertainty will always remain. In other words, experiments in social policy can be extremely instructive, especially when compared to the ignorance in which so many social programs are debated. But they are only helpful; they can never be truly conclusive.

Second, if exemplary true and quasi-experiments cannot produce definitive results, it follows that findings from studies using weaker designs may be worse than useless. On one hand, the research may be so

obviously flawed that scientific methods themselves may inappropriately take the rap. The early evaluations of Head Start, for example, helped stimulate across the board skepticism of "positivist"⁹ methods. On the other hand, misleading findings may be seized upon, perhaps for the wrong reasons, and influence social policy. A recent study claiming to show that parochial schools were more effective than public schools (Coleman et al., 1983), for instance, is both badly flawed (Rossi, 1984) and widely quoted. It follows, therefore, that powerful research designs and data analysis plans are essential and that in addition, resources must be provided to capitalize on them. A research design that founders for want of sufficient support would be better off left on the drawing board.

Third, even exemplary social policy experiments, however, do not necessarily lead to appropriate social policy. Despite state-of-the-art research methods, the findings may be wrong, misconstrued, or simply ignored. For example, there is no credible evidence for the deterrent effect of capital punishment (Blumstein et al., 1980), but that has not kept public officials from claiming that the death penalty deters. Likewise, strong evidence that price is an effective water conservation tool (Berk et al., 1981) has had little impact on continuing low prices for agricultural users, clearly encouraging waste (and, incidentally acting as a subsidy to agribusiness). We raise these difficulties in part to provide a realistic picture of the current scene, but also to support later arguments for efforts that might improve matters.

Fourth, political fashions come and go and with them concerns for particular social problems. One result is haphazard social policy research programs. Social policy experiments need to be planned carefully with an eye toward *packages* of studies. Single experiments hastily initiated to answer the questions of the moment may on occasion provide useful information, but are unlikely to produce the *cumulative* body of knowledge required for sound policy decisions. For example, a recent experiment showing that under certain conditions, arrests deter incidents of wife battery (Sherman and Berk, 1984) should really have been undertaken as part of a general research program on law enforcement responses to family problems. We have an isolated finding when we should have a growing body of knowledge. In response, means will be suggested later to provide some coherence to social policy experiments.

Finally, all three experiments made little use of social science theory and in turn have been essentially ignored by the basic research

community. It is likely that each study would have been significantly enriched and strengthened had the treatments been more directly based on what social science has to say about the impact of monetary incentives and employment. It is also likely that basic research on crime is all the poorer as a consequence of ignoring what was found.

IV: SOCIAL POLICY EXPERIMENTS MORE FORMALLY DEFINED

WHAT IS SOCIAL POLICY EXPERIMENTATION?

A social policy experiment is here defined as *a theoretically based effort to introduce social change in a way that permits discerning the net effect of the change on social outcomes*. Thus, an explicit and clearly articulated alteration in policy or practice is expressly linked to the need to determine what consequences follow, after taking into account other things that might affect those consequences.

We envision two major purposes for social policy experiments: first and foremost to inform policy decisions and, second, to advance social science theory and knowledge. In the past, these two enterprises have been artificially separated to the detriment of both. It is, in fact, hard to imagine good social policy not capitalizing on the accumulated wisdom of the social sciences, and equally hard to imagine changes in social policy from which social scientists could not learn. But our point here is that both the applied and basic research audiences can benefit from social policy experiments. The TARP experiment, for instance, had important implications for microeconomic theory and for whether transfer payments delivered to ex-offenders under different taxing schemes were an effective way to reduce crime.

The form of social policy experimentation that we advocate here has three distinctive features. First, it must exploit social science theory. Many policy interventions that are seriously proposed may seem sensible on their face, but are rarely thought out carefully, and rarely capitalize on theories from psychology, economics, sociology, and political science. This is no small loss, because social science theory could make important contributions when an intervention and research design are formulated. A serious scrutiny of social science theory would

also alert researchers to potential implementation problems. For example, in developing a water conservation program and an appropriate evaluation design, it would be essential to capitalize on microeconomic theory, asserting that for water to be used efficiently, the price (per unit) must equal the marginal cost. Efficient allocations will not follow, in contrast, if the price consumers pay per unit is equal to the average cost (which is common practice).

At the same time, what we mean by theory is quite elastic. It ranges from the lay "let's try this," to propositions deductively derived from first principles. We take this broad stance in part because we could not agree (and we are hardly alone) about what constitutes a theory, and in part because it seems silly to exclude lay understandings of how things work simply because social science certification is lacking. The key point is that social theories are meant to explain *how* a particular social process works, and we are stressing the need to think things through carefully, exploiting all available information. In any case, we certainly do not want to get bogged down in a long and abstract discussion of the nature of theory, especially when it is unnecessary for our purposes.

Our requirement that existing theory be exploited excludes ad hoc efforts to "tinker with the system," and policy changes that fail to recognize existing knowledge. So, for example, we regard tinkering as a legitimate and important activity for public program managers and program staff. We could easily imagine, for instance, a teacher intuitively altering his or her classroom technique in response to the mix of pupils. But such efforts fall outside our purview unless the intervention is carefully considered, bringing extant knowledge to bear.

We also exclude all forecasts if they are not backed by a description of the underlying causal mechanisms. One could easily imagine, for example, a company forecasting demand for its products by simple extrapolation from past demand. Likewise, one might forecast the unemployment rate by extrapolating from historical (e.g., seasonal) patterns. Although both forecasts may be quite accurate and useful, they are not theoretical statements; they fail to address how things work. On a more subtle level, a forecast might draw on "leading indicators." Yet, unless there is an accompanying explanation of how the leading indicators are causally linked to the outcome, no theoretical statements are being made.

A second distinctive feature of our perspective on social policy experimentation is that the social changes or interventions of interest

are necessarily policy mutable. They include changes in law (e.g., from indeterminant to determinant prison sentences), corporate policy (e.g., the introduction of workers councils), administrative regulations (e.g., certification requirements for food stamps), and other instruments important enough to justify a planned effort to understand their effects in the least equivocal manner. They would exclude changes that for all practical purposes were not subject to manipulation: natural disasters (e.g., earthquakes), charismatic social movements (e.g., the rise of the New Right), historical events (e.g., the election of Ronald Reagan), and the like.

To illustrate more thoroughly, the age mix of the population surely has an impact on crime, school crowding, and the demand for certain kinds of consumer goods, but in the short run at least, and under current social arrangements in the United States, the mix of young to old people in the population is not subject to manipulation. Likewise, one might be interested in the impact of race on income, but it is not possible to alter someone's race to determine whether, for example, people assigned to "black" do worse than people assigned to "white."¹⁰

It is possible, of course, to explore the relationships between age mix and crime, school crowding, or consumer demand with nonexperimental methods. Similarly, one could examine the relationship between race and income with nonexperimental methods. Much, in principle, could be learned. However, we are advocating experimental approaches because of a general consensus that causal inferences from nonexperimental methods are typically inferior to those obtained from experimental methods (Boruch and Reiken, 1974; Cook and Campbell, 1979; Holland and Rubin, 1984; Rosenbaum and Rubin, 1983; Rossi and Freeman, 1982; Pratt and Schlaifer, 1984; Dawid, 1984; Rosenbaum and Rubin, 1984). Indeed, Holland and Rubin would consign all nonexperimental social research to mere description.¹¹

Lest we be misunderstood, we are *not* arguing that the only useful applied social research is based on experimentation. There are times when causal inferences are not at issue, such as when we try to estimate the number of families in the United States living below the poverty level. We are then only interested in some quite important "mere description."

There will be other times when causal inferences are desirable, but when experimentation is impractical. One might then settle for non-

experimental studies as a second best alternative. Many would argue, for instance, that it would be unethical to undertake an experiment on the psychological effects of having an abortion. Even if a group of pregnant women could be found who would volunteer for random assignment to treatment and control groups, the fetus' interests are not considered. Under these circumstances, one might proceed with a nonexperimental design in which women who had an abortion were first matched with women who had not had an abortion (see for example, Rosenbaum and Rosen, 1983, for a discussion of matching). Insofar as the experimentals and controls were matched on all of the important variables that might otherwise explain any differences in psychological functioning (e.g., age, religion, marital status, income), plausible causal inferences about the *net* impact of an abortion might be made. These might serve as the best policy-relevant results available on which programmatic decisions could be made or as formal hypotheses from which more powerful studies could be designed.

Finally, there will be still other times when alternative attributes of sound, applied research are more important than making the most credible causal inferences. Suppose, for example, that one were interested in the impact of avoidance conditioning on alcohol abuse. However, the individuals who are available for random assignment to avoidance conditioning (i.e., punishing instances of the undesirable behavior) are under a court order to undergo immediate treatment (e.g., because of a drunk-driving conviction). It would be very difficult, therefore, to generalize one's findings to other kinds of persons in need, and one might choose to undertake a study based on matching (or statistical adjustments) if a more representative sample could be obtained.

We also do not mean to ignore that role of applied research that *supports* good social experiments. Thus, it often is essential to do applied studies determining the nature and extent of a particular social problem before a program is designed (i.e., "needs assessments"). A recent instance is investigations into whether there are a large number of Americans who were malnourished. Similarly, it is often desirable to determine whether a given social program is operating as designed. Here, under the label of "monitoring," *ethnographic methods* are essential. Even a casual inspection of many of the projects launched under the Model Cities Program of the 1960s, for instance, revealed that

a number of programs for which funds were allocated were never actually implemented. Hence, a failure to find any program impact could often be explained by an absence of a program rather than by program ineffectiveness. And in a similar fashion, analyses of cost-effectiveness are typically helpful, because so many programs should be justified in monetary terms. For example, given the costs of incarceration, it is not at all clear that recent legislation requiring longer prison terms is cost-effective, even if (and this has not been demonstrated) longer prison terms reduce crime. In short, there is nothing in our perspective that denies the importance of such activities as needs assessments, monitoring, or benefit-cost analysis. We view them, however, as being in service of determining the net impact of policy innovations; we do not see them as ends in themselves.

The third distinctive characteristic of our approach is that we regard as crucial the requirement that when experiments are attempted, the government or others provide resources for estimating the impact of the changes in the program or policy, before the changes are introduced. This prior commitment distinguishes social policy experiments from *ex post facto* analyses. It also permits the design of research that leads to less debatable, more precise, and more reliable information about the outcomes of the intervention.

We briefly mentioned earlier, for example, an experiment undertaken to determine which of three police responses to wife battery most reduced the number of postintervention wife-battery incidents. In the experiment, arresting the offender proved more effective than ordering the offender from the premises for several hours, or trying to mediate the dispute. Although the study relied on random assignment of households to one of the three treatments, allowance had to be made for instances in which the separation and mediation interventions had to be "upgraded" to an arrest (on a nonrandom basis). If an offender refused to leave the premises when ordered to do so, for instance, the officer had to be given the option of making an arrest. This in turn meant that data had to be collected on (1) the assigned treatment, (2) the treatment actually delivered, and (3) variables that might affect the shift from separation or mediation to arrest. It is extremely unlikely that such information would have been recorded had not adequate resources for the study been provided; police do not typically indicate in much detail (if at all) why a particular action is taken. And under those circumstance, a far weaker

analysis of the data would have resulted. In short, we envision social policy experiments as requiring a commitment to research backed by the necessary resources. This excludes from our definition such programs as Jesse Jackson's Push for Excellence, because meaningful resources have not been allocated in advance to determine the program's impact. Excluded for the same reason are activities such as "scientific jury selection" (which uses social science technology to help in the selection of jury panels).

WHY SHOULD WE DO SOCIAL POLICY EXPERIMENTS?

The principal beneficiary of social policy experiments is society itself. It should not have to be argued that with more reliable information about the impact of policy alternatives, the opportunity exists for more effective public policy. We are well aware that the opportunity may not translate into real benefits; the information may be ignored or even misused. Better information is surely a necessary prerequisite for improved responses to social problems.

To illustrate, educational foundations have relied on randomized field experiments to discover the beneficial effects of *Sesame Street*, *The Electric Company*, and other programs, and to justify continued support. Yet, randomized tests of the television show *Feeling Good* revealed that despite good intentions and glowing rhetoric, some forms of television-based health education for poor families simply do not work. The negative findings were used to terminate a program whose costs would have far exceeded any benefits, and to reallocate scarce resources to more effective approaches to health education.

In addition, there are often costs involved in *not* experimenting. Clearly, early Comprehensive Employment and Training Act (CETA) programs produced no credible evidence about their impact, and as a consequence, public debate was confused, money was poorly spent, and human expectations were dreadfully misdirected. To take a more recent example, many millions of dollars are now being spent introducing "computer literacy" into schools. All kinds of grand claims are being made. We are told by some, for instance, that learning how to program

will teach the kind of conceptual skills that children need to master algebra. Or, we are told that with computer literacy will come better coping abilities needed in the world of the twenty-first century, or that people comfortable with computers will more easily find employment. But none of these claims has been subjected to randomized field experiments. It may be that most of the claims are false, in which case a lot of expectations have been inappropriately raised and enormous amounts of money wasted. It may be that there are some benefits from the introduction of computers into schools, but that the same money spent in more traditional ways would be at least as effective. Or it may be that a number of benefits follow, but as a result, students from less richly endowed schools will fall still farther behind their more fortunate peers. Perhaps some ways of using computers are effective although others are not. In short, it may be very costly not to know; a failure to experiment may well incur substantial costs.

There are also a number of somewhat more subtle benefits from social policy experimentation. First, there may be high political costs in launching untested programs that fail. If a substantial investment is made in responses to a particular social problem, and the social problem persists, political leaders may throw up their hands and conclude that the problem is intractable. An excellent recent example is unemployment. Over the last decade, as the unemployment rate has grown in the face of many remedial efforts, politicians have gradually redefined to higher and higher levels what is called the "natural" rate of unemployment. Alternatively, one might argue that the remedial efforts were not properly designed, tested, and implemented. The CETA program, for instance, was at best terribly administered.

Second, social policy experiments can serve as inspiration. An experiment can be undertaken before there is a full political commitment even to seeking a cure for the problem. Finding that something works may inspire a legislator to "discover" the problem by providing a particular course of action to pursue; a legislator with a hammer will likely look for a nail. For example, the experiment described earlier on police interventions in wife-battery incidents (which was very widely publicized) has led legislators in a number of states to discover the problem of law enforcement responses to family violence.

Third, experiments catch the eye. They seem to speak for themselves as a self-contained and comprehensible response to a particular social problem. They give the appearance that someone is trying to do something. Other forms of applied social research are less readily interpreted in this manner. There were a great many correlational studies of the relationship between cholesterol and heart disease before the recent experiment; it was the experiment that made the cover of *Time* magazine.

Fourth, social policy experiment that are done well will give increased legitimacy to the scientific enterprise at a time of considerable public skepticism. From the Fundamentalist attacks on the theory of evolution to quite proper concerns about the environmental implications of genetic engineering, science has been on the defensive for a number of years. Good social policy experiments may help to counter this trend.

Fifth, good social experiments educate the public about the scientific method and increase public understanding about the sorts of issues that must be addressed to provide compelling answers to factual questions. The senior author of this report, for example, was recently presenting some nonexperimental results on the causes of prison crowding to a meeting of a local Optimists Club, and was quite properly grilled about the credibility of his causal inferences. Indeed, one of the members (an owner of a nursery) developed an argument that basically captured the ideas behind regression to the mean. Visible social policy experiments may well foster such sophistication.

Finally, social policy experiments permit adventurousness tempered with caution. Properly presented, social policy experiments should appeal to both the audacious and the deliberate. Creative innovations of all kinds can be implemented on a small scale, and their effects examined. Social policy experiments can be progressive in the interventions tried and yet conservative in the speed with which full implementation may be undertaken.

In the face of these and other benefits, the case against social policy experimentation seems quite weak. Most of the criticism addresses much larger issues beyond social policy experimentation per se. Thus, claims remain unchallenged that social policy experiments are inherently "conservative" in that the innovations tried are modest in scope and

impact, leaving the basic “system” intact. By and large, this is an accurate portrayal of social policy experiments, but it also describes virtually all social change mechanisms available within “democratic” societies. No one is claiming that social policy experimentation is a ready means to social revolution.

Others claim that because social policy experiments are typically initiated by those in power, social experiments will only help the powerful at the expense of the powerless. Whether or not this is true, it is not an inherent feature of social policy experiments. To run social policy experiments does require access to funding and enough political power to secure cooperation from public officials, but large numbers of diverse groups have succeeded in the past. For example, effective applied research has been undertaken by labor unions, consumer groups, environmental groups, and—perhaps most recently and visibly—by an informal coalition of scientists who demonstrated that a “nuclear winter” might well result from a nuclear exchange. There is no reason why social policy experiments cannot in a similar fashion respond to concerns raised by a wide variety of interests.

Still others claim that social policy experiments undermine the democratic process by putting political and value questions in the hands of technocrats. Although we certainly do not claim that there is any simple distinction between fact and value, social policy experiments are meant to produce knowledge based on scientific methods. And there is nothing in those methods that need imperialize activities in the political arena. For example, the decision about what sorts of social problems are important enough to address through experiments is primarily a political question. Likewise, decisions about what sorts of experimental effects are important enough to measure is primarily a political decision. Social policy experiments do not inform such judgments.

In the end, the major objections to social policy experiments are practical and particular: Is a given experiment too costly? Does a given experiment violate important ethical principles? Can a given experiment deliver the information required in the time available? Can a given experiment be properly implemented in the field? These are all important concerns, and occasionally there will be insurmountable problems, but by and large, the usefulness and feasibility of social policy

experiments has been amply demonstrated. What is needed, therefore, are means to capitalize on the promise of social policy experiments. In the pages ahead we will suggest some ways in which this might be done.

V: A PROGRAM FOR SOCIAL POLICY EXPERIMENTATION— AN ORGANIZING STRUCTURE FOR SOCIAL POLICY EXPERIMENTATION

We believe that wise social experimentation is most likely to occur if an organization is created to serve as orchestrator. Such an organization can help promote experiments and help avoid a piecemeal approach that will waste scarce resources. A possible organizing structure would be one or more centers for social policy experimentation. There are excellent examples in other disciplines of centers on which to draw. The Salk Institute, National Center for Atmospheric Research, and the Institute for Theoretical Physics are three instances with which we are acquainted.

At this point, however, we do *not* for several reasons favor establishing any centers for social policy experimentation. First, centers are typically very costly to start and maintain. Second, even if money were available to start a center, that money might well be better spent in other ways; the opportunity costs would be great. Third, centers too often ossify, as researchers, administrators, and staff become invested in the center *per se* and not its activities. Fourth, centers concentrate massive resources in a few locations and in a small number of hands. The possibility of large and relatively *indivisible* resources almost inevitably triggers a fierce and often self-destructive scramble among prospective suitors. In the end, a few win big and most lose big. This is not the atmosphere in which cooperative scientific endeavors thrive. In short, all-or-nothing competitions are sometimes useful, but we see no compelling advantages for them here. Fifth and finally, we think that there is an alternative structure that will accomplish much the same goal at less cost and with fewer liabilities.

We propose the establishment of a committee on social policy experimentation (or some similar name) whose main purpose would be

to stimulate the use of social policy experiments addressing pressing social problems. The committee would have about 15 members knowledgeable in experimentation who would meet frequently and regularly; perhaps once a month for two days year round. Appointments to be committee would be for terms of 3 to 5 years.

The committee would have an executive officer in charge of administration on a day-to-day basis: providing agendas for the monthly meetings, overseeing support services, and the like. The appointment would be for a maximum of five years. There would also be associates of the committee who would be more junior persons, serving essentially as committee staff. Finally, there would be support personnel for clerical work and bookkeeping.

At the monthly meetings, three broad tasks would be addressed. First, the committee would *develop and advocate policy-fostering experiments and other related research activities, programs and projects*. These would include such actions as (1) identifying and suggesting needed technical improvement in research methods, (2) identifying and suggesting measures needed to augment the supply of social scientists and others trained in the requisite skills for social policy experiments, and (3) identifying and suggesting enabling legislation needed to support experimental approaches to social problem solutions. For example, the committee might decide that experimental approaches developed by engineers to evaluate the reliability of complicated pieces of machinery had potential applications in the evaluation of social programs. One action might be to outline and then “peddle” to funders a postdoctoral program for social scientist who wanted to learn about engineering quality control techniques.

Second, drawing on a variety of sources, the committee *would identify and suggest social issues needing new policies and programs*. When such problem areas were found—and assuming they were sufficiently serious, well defined, and tractable—a special, *ad hoc subcommittee would be formed (recruited from outside the national committee) to explore experimental and other research opportunities*. For example, responding to recent publicity, a subcommittee might be formed around the problem drunk driving. That subcommittee might develop a set of experimental designs to evaluate the impact of recent drunk-driving laws and then “shop” for funding with which to get the

research done. Alternatively, they might first seek funders and then in collaboration, design and implement the study.

Finally, the committee would *recommend to foundations, government agencies, and private enterprise substantive areas in which programs of experimentation offer promise*. Examples might include tax compliance, incentives systems to reduce acid rain, the needs of the homeless, and the deterrent effect of incarceration.

To make the functioning of the committee more concrete, consider the following example. Technological obsolescence is a recurrent problem in certain industries. Companies such as Exxon, DuPont, and Boeing as well as service industries and government agencies are routinely affected. Suppose that technical obsolescence is selected by the committee as one of several problems that need to be pursued. It would be apparent that a number of different consequences are involved: productivity, plant closings, structural unemployment, worker safety, balance of trade with foreign countries, and many others.

A subcommittee would then be constituted to address macrolevel issues (e.g., balance of trade) and microlevel issues (e.g., worker retraining), pulling together what was known about the causes and consequences of technological obsolescence. At this point, the abilities of an expert and ecumenical subcommittee would be critical as the subcommittee would have to decide what kinds of experiments at what level (micro or macro) would best advance the policy-formulation process (as well as scientific progress).

Assuming that such experiments could be described, the next step would be to facilitate the desired field research. Options might include the following (other possible activities are described more generally later):

- (1) contract or grant-funded experiments to assay effectiveness of theory-based "solutions" to one or more of the most pressing problems;
- (2) contract or grant-funded experiments *adjoined* to federal/state demonstration projects to generate theory-related evidence about whether particular projects work, why they work, and whether a positive benefit-cost ratio results;
- (3) contract or grant-funded reanalyses of data from earlier experiments for what could be learned about responses to technical obsolescence;

- (4) contract or grant-funded efforts to solve methodological problems preventing sound experimentation on the problem; and
- (5) "seed money" and "matching funds," efforts to get any of the four options funded by other private or public sources.

AVOIDING PITFALLS

Given the activities and structure proposed for the committee, there are clearly a number of pitfalls to be avoided. First, the danger of ossification exists; long-term participation may make committee members stale, tired, or even lazy. In response, we recommend that no committee appointments be for more than three years, with each appointment made for one year at a time.

Second, there is the risk of special pleading by committee members. In response, we recommend that no committee members can apply for or receive grants during their tenure on the committee from funding sources used by the committee for social policy experiments; basically they must not be permitted to use the committee to line their own research pockets.

Third, committees of this general type can founder on lack of commitment; members just go through the motions. In response, we recommend that a reasonable stipend be paid to committee members for their work. Alternatively, released time from teaching or administration might be purchased.

Fourth, it is possible for such committees to become dominated by their staffs. Committee meetings may be little more than rubber stamps for work done by full-time, paid staff members. In partial response, we recommend that staff appointments be for a single term of perhaps no more than five years or treated as fellowships of even shorter duration.

Fifth, there is the risk of the committee becoming too cozy with a particular discipline or point of view. To encourage heterogeneity, we recommend that "slots" for different disciplines be established. There should also be a good mix of individuals from different professional settings (academe, government agencies, private research firms, etc.), who are at different stages of their careers.

Sixth, there is a danger that the committee will become too "political" and lose its credibility. That is, the committee is surely a vehicle for

enhancing social policy experimentation, but must be seen as reasonably evenhanded on a variety of other dimensions. It must try to avoid, for example, being identified with either liberal or conservative camps (e.g., The Brookings Institute and the Hoover Institute). We recommend, therefore, that the committee not function as a lobbying group with Congress or other official decision-making bodies and that the committee avoid making public pronouncements on the issues of the day, except issues relating to research and scientific methods themselves. We are not suggesting, of course, that social policy experiments or the committee could ever be "value free." Values of various sorts must necessarily be embodied in whatever actions are taken. We are arguing however, that the committee avoid traditional political activities and that the committee respond fairly to a wide range of value positions.

Finally, there is the possibility that the committee will not function as planned and either gradually take on a less desirable agenda or simply perform in an ineffective manner (compared to other ways to facilitate social policy experimentation). Hence, we recommend that in addition to yearly auditing and substantive reviews, there be every five years a thorough assessment by an outside panel of academics, public officials, and others knowledgeable in social policy experimentation. The committee itself is an experiment and needs to be evaluated.

In summary, we are proposing the establishment of a committee to facilitate and shape at a national level the use of field experiments to inform public policy. In general terms, therefore, there are ample precedents; the Social Science Research Council and various committees of the National Academy of Sciences are examples. However, our proposed committee is in many ways unique.

To begin, it has a relatively narrow mission. There is no intent to directly influence a wide variety of social science activity, much as the Social Science Research Council does. The more specific focus will prevent the committee's overriding agenda from becoming diluted in a host of other concerns about the social sciences or science more generally. Likewise, the committee is not an appendage to some larger collection of scientists. Hence, the committee is free to establish its own agenda unencumbered by unrelated issues, historical baggage, or ongoing commitments. This freedom is all the more important given the unusual mandate explicitly combining scientific and policy concerns.

ACTIVITIES THE COMMITTEE MIGHT FOSTER OR UNDERTAKE

With the committee's structure described, we can turn in more detail to the kinds of activities that might be undertaken by the committee or with its encouragement and support. What follows is a list of possibilities varying in size, likely impact, and cost.

Direct Involvement in Social Policy Experiments

(1) *Funding particular social policy experiments.* We have already tried to make the case for social policy experiments and hope that nothing of major importance still needs to be said.

(2) *Funding replications of social policy experiments.* Over the past 20 years social policy experiments have been funded from time to time, and many of these have been very instructive. Replications are much less commonly supported. Yet, replication lies at the heart of generalized knowledge.

(3) *Funding for particular secondary analyses.* Social policy experiments often produce large data sets that are incompletely analyzed by the original investigators. In addition, there are often alternative data analytic approaches, occasionally leading to alternative results, not all of which are explored by the original investigators. It is often useful, therefore, to support a set of analyses of the same data undertaken by several independent groups of researchers.

(4) *Funding particular meta-analyses.* A meta-analysis is basically an attempt to draw some overall conclusions from a series of social policy experiments on a given topic (e.g., the impact of deinstitutionalization on mental patients). Summaries of the literature have been undertaken as long as there have been literatures to summarize. Recent work in meta-analysis has tried to apply more rigorous, statistical procedures for formally aggregating results over the set of experiments, and has advanced understanding about what works and the extent to which it works.

Providing Resources for Social Policy Experiments

(5) *Support for field laboratories in social policy experimentation.* In the environmental sciences, the National Science Foundation supports the maintenance of several large research sites (e.g., a desert setting, a coastal setting) in which environmental scientists can conduct observational studies and experiments. In such sites, it is possible, for example, to dam a stream and see what ecological consequences follow. One could imagine establishing ongoing research sites for social policy experimentation. These might be designated as individuals, households, geographical areas (e.g., neighborhoods), or institutions (e.g., a particular police department, a particular school system), *voluntarily* making a relatively long-term commitment to participation in a series of social policy experiments. There are a number of advantages to this approach, including the prospect of reduced costs, better quality control, and especially rich data sets.

(6) *Endowed chairs in social policy experimentation.* Endowed chairs would give added stature to social policy experimentation and applied social science research more generally. As noted earlier, applied social science research typically carries low prestige activity in academic circles. As a result, talented social scientists often avoid applied work and their students primarily pursue basic research. The prestige of endowed chairs may help to right the balance. In addition, endowed chairs will allow energetic applied researchers to be devoted full time to social policy experiments.

(7) *Subsidized publications in social policy experimentation.* Many important reports of social policy experiments go unpublished because they are thought by publishers to be unprofitable. Whether they are or not, a series of subsidized publications (e.g., a monograph series) would help to give stature to the field and disseminate important findings.

(8) *"MacArthur awards" for particular social policy experiments.* One might designate several "MacArthur awards" per year for someone active in social policy experimentation. Besides allowing that individual to pursue full-time experimentation, the award would add legitimacy to the field.

(9) *“Blank check” research awards for some period of time.* These would be much like the “MacArthur award” but would be given solely to support research activities in social policy experimentation (of the recipient’s choosing.)

(10) *Support for making social policy experiments a greater part of the political process.* Examples include funding trips to testify at legislative hearings, funding a newsletter on social policy experiments, holding conferences on social policy experimentation for the news media, funding meetings between social policy experimenters and legislative aides and others. It includes establishing special, ongoing seminars to enhance *regular productive exchanges between experimenters and legislators willing to test their ideas.*

(11) *Conferences for academics on social policy experimentation.* There may be too much “conferencing” among academics already, but maybe not.

(12) *Annual awards for outstanding work on social policy experimentation.* One might support something like an R. A. Fisher Award to be given each year by the American Statistical Association or the Evaluation Research Society for outstanding contributions to social policy experimentation. The award might well include a substantial honorarium.

Training and Retooling

(13) *Paid leaves or sabbaticals for university faculty who want to train or retool in social policy experimentation.* The recent flurry of interest in applied social research has demonstrated that there are a number of academics who might pursue more applied careers if they could obtain the necessary technical skills. One might, for example, have a postdoctoral program for academics five years or more beyond the Ph.D.

(14) *Paid leaves or sabbaticals for government officials, business executives, and others who want to train or retool in social policy experimentation.* Many government employees, among others, would seek training in social policy experimentation if they could afford to do so. Several years ago, summer training programs in evaluation research

at Northwestern University and the University of Massachusetts routinely drew a considerable number of government employees.

(15) *Postdoctoral training programs.* Although no labor market surveys have been done, impressionistic evidence suggests that there are many jobs for individuals trained in social policy experimentation. For example, postdoctoral programs in evaluation research at UCLA and Northwestern University always place their students. We also have had success placing our students, many of whom have applied interests. In addition, it is entirely possible that with more people trained in social policy experimentation, interest in social policy experimentation will grow (a kind of supply-side effect), and grow in a productive manner. At the very least, the legitimacy of the social policy experimentation would be enhanced.

(16) *Pre-doctoral training programs.* The arguments for pre-doctoral training programs are much the same as those for post-doctoral training programs, except that one is likely to be educating not just individuals who will do social policy experiments, but individuals who will have to understand them and put their findings into practice. That is, "consumers" of social policy experiments will be trained as well: corporate executives, elected officials, government administrators, and others.

(17) *Retooling programs for government personnel.* There is no reason why training programs in social policy experimentation should exclude people who already have established careers. One could imagine a range of retooling efforts lasting from a weekend to several years. It is entirely conceivable that many government agencies would pay for such training programs. Summer programs on applied research have in fact been reasonably successful in the past.

(18) *Retooling programs for business personnel.* The arguments for retooling that apply to government personnel apply equally to business personnel. Many firms have active applied research programs and in fact undertake social policy experiments as we have defined them (e.g., on worker productivity). They also have to know how to interpret social policy experiments undertaken by others (e.g., on vocational training programs). Like government agencies, businesses might well pay for training programs in social policy experimentation.

(19) *Support for the development of curricula for colleges in social policy experimentation.* To the best of our knowledge, there are very few courses in social policy experimentation currently being offered to

undergraduates. There has not even been very much systematic thought about what such courses should contain. One might support individuals or groups (or hold a conference or two) wanting to improve existing curricula (e.g., by “buying back” release time from their other duties).

In short, there are clearly a number of activities in support of social policy experimentation that our proposed committee could undertake or directly encourage. The problem will not be finding useful things to do. The problem will be to select among a rich menu of opportunities.

SOME BRIEF OBSERVATIONS ON GETTING FROM HERE TO THERE

Even if the views expressed in this position paper are persuasive, many steps remain before our committee on social policy experimentation could be launched. Perhaps most important, we have not considered the various ways one might get from the current situation to a more organized and visible presence for social policy experiments.

On the surface, the best approach might be from the top down. One or more funding organizations could provide the necessary support, and within a relatively short time the national committee would in principle be functioning. Certainly, the costs of the committee is not a significant obstacle relative to other investments in the social and natural sciences.

However, it seems to us that such a scenario is unlikely. As a practical matter, there is fierce competition for existing resources, and the game is sometimes very close to zero sum. In addition, there is in applied social research—just as in any enterprise—a considerable commitment to the status quo. We are proposing substantial change. Finally, for reasons discussed earlier, social experimentation has been viewed with some hostility by the basic research community and with some skepticism by policymakers. In short, whatever the merits of our national committee, there will no doubt be significant opposition.

Therefore, a bottom-up strategy may be more productive, and several complementary approaches are possible. Thus, if the national scale is too grand, significant progress might be made at the state or even local level. Likewise, one might begin within a particular policy area or in concert with a single government agency. For example, the Attorney General of California recently proposed collaborative undertaking in applied social research between the state’s Bureau of Criminal Statistics

and the University of California. Central to this effort would be a governing committee much like our proposed national committee. And perhaps most important, social policy experiments would figure significantly in the collaboration. Should this effort prove productive, similar arrangements might be established in a number of agencies and ultimately at the national level.

Another bottom-up approach might begin with one or more of the specific initiatives briefly described above, with the goal of fostering a community of interest in social policy experiments. For example, a foundation or government agency might be persuaded to fund a package of social policy experiments to be initiated over a five-year period. These experiments would be billed as exemplary efforts tackling important social problems using state-of-the art methods and theory. The key element, however, would be regular meetings among individuals associated with the experiments to exchange views and experiences. A wide variety of disciplines and substantive interests would be represented, and over time an influential constituency for social policy experiments could develop. It might then be a short step to a national committee on social policy experimentation.

A CLOSING EXHORTATION

The scientific arguments for social policy experiments are well-known and widely accepted when causal inferences are essential to the policy formation process. It is also apparent, that the practical and ethical obstacles to social policy experiments have been vastly overrated. Finally, useful social policy experiments can be funded for modest sums and even rather dear social policy experiments can be highly cost-effective. Thus, an intellectual case against social policy experiments is difficult to construct. It is time to get on with the operational problem of integrating social policy experiments into the policy formation process.

NOTES

1. We use the term "social policy experiments" rather than "social experiments" to emphasize the explicit links to the policymaking process and to negate some of the "Strangelovean" connotations associated with the term "social experiments."

2. This rendering of the income maintenance experiments is vastly simplified for purposes of exposition. The same holds for the other experiments summarized in this section. In each case, a full discussion can be found in cited material. Nevertheless, it is perhaps worth mentioning that the later income maintenance experiments undertaken in Denver and Seattle found modest work disincentive effects across the board.

3. One does not have to randomly assign subjects. If one has a longitudinal design, one can also randomly assign *when* the intervention occurs.

4. For example, the *Evaluation Review*, which is the journal most widely subscribed to in the field of evaluation research (which is built around social experimentation), is only 7 years old. The Evaluation Research Society is about the same age.

5. In practice, anticipating whether an experiment will be cost-effective is very complicated (Mosteller and Weinstein, 1981). For example, whether an experiment will be cost-effective depends in part on what the findings will be. But if these were known, there would be no need to do the research.

6. However, there was evidence that the unemployment benefits helped ex-offenders find better (i.e., higher-paying) jobs, just as anticipated.

7. The legislation required, for example, both an impact assessment on a number of outcomes and monitoring of how the program was implemented. Such specificity is very rare and shows how social policy experimentation as a method was already affecting legislative thinking (at least in some circles).

8. It is not clear whether the failure to pass an appropriation was due to an oversight, a belief that the Department of Corrections had the necessary resources, or political opposition. One could imagine, for example, a decision not to push for an appropriation when much of the bill's justification rested on saving tax dollars.

9. For those who care, positivism is an outdated philosophy of science that no one familiar with the issues takes seriously any longer. Yet, positivism has been adopted as a rallying cry by many of those who object to quantification, statistics, or scientific methods more generally. Although quantification, statistics, and scientific methods are hardly flawless and do overlook or fail to capitalize on other ways of "knowing," no one has yet proposed in any detail what the alternative might be. Serious statements that have appeared are either broadly philosophical (and, therefore, offer little on how concretely to proceed) or are more properly thought of as complements to experimental approaches.

10. And it is not clear what one learns from cross-sectional comparisons. It is hard to accept that comparing the incomes of white and black people, other things equal, will tell you much about what would happen if race were manipulated at random; imagine waking up one morning with your race changed.

11. According to Holland and Rubin, unless the treatment can be manipulated, there can be no experiment; indeed, the notion of cause does not apply. Thus, one cannot meaningfully ask, as do Pratt and Schlaifer (1984), whether there would be more snow in Denver if the Rocky Mountains were lower. However, we are *not* limiting experiments to instances in which treatments are assigned randomly. We *include* quasi-experiments.

REFERENCES

- AIGNER, D. J. and J. A. HAUSMAN (1980) "Correcting for truncation bias in the analysis of experiments in time-of-day pricing of electricity." *Bell Journal* 35: 405.

- BENNETT, C. A. and A. A. LUMSDAINE (1975) *Evaluation and Experimentation*. New York: Academic.
- BERK, R. A. (1981) "On the compatibility of applied and basic research: an effort in marriage counseling." *Amer. Sociologist* 4: 204.
- R. A. and D. RAUMA (1983) "Capitalizing on nonrandom assignment to treatments: a regression discontinuity evaluation of a crime-control program." *J. of the Amer. Statistical Assn.* 78: 21.
- BERK, R. A., C. J. LaCIVITA K. SREDL, and T. F. COOLEY (1981) *Water Shortage: Lessons in Conservation from the Great California Drought, 1976-1977*. Cambridge: Abt.
- BLUMSTEIN, A., J. COHEN, and D. NAGIN [Eds.] (1978) *Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates*. Washington, DC: National Research Council.
- BORUCH, R. F. and W. SHADISH (1982) "Design issues in community intervention research," in E. Seidman (ed.) *Handbook of Community Intervention Research*. Beverly Hills, CA: Sage.
- BORUCH, R. F., A. J. McSWEENEY, and E. J. SODERSTROM (1978) "Randomized field experiments for program planning, development and evaluation: an illustrative bibliography." *Evaluation Q.* 2: 655.
- BORUCH, R. F., P. ANDERSON, D. RINDSKOPF, I. AMIDJAYA, and D. JANSON (1979) "Randomized experiments for planning and evaluating local programs: a summary on appropriateness and feasibility." *Public Administration Rev.* 39: 36.
- CAMPBELL, D. T. (1969) "Reforms as experiments." *Amer. Psychologist* 24: 409.
- COLEMAN, J. S., T. HOFFER, and S. KILGORE (1982) *High School Achievement: Public, Catholic, and Private Schools Compared*. New York: Basic Books.
- COOK, T. D. and D. T. CAMPBELL (1979) *Quasi-Experimentation: Design and Analysis in Field Settings*. Skokie, IL: Rand McNally.
- CRONBACH, L. J. (1983) *Designing Evaluations of Educational and Social Programs*. San Francisco: Jossey-Bass.
- DAWID, A. P. (1984) "Comment on the nature and discovery of structure." *J. of the Amer. Statistical Assn.* 79: 22.
- DIAZ-GUERRERO, R., I. REYES-LAGUNES, D. B. WITZE, and W. H. HOLTZMAN (1976) "Plaza Sesamo in Mexico." *J. of Communications* 26: 145.
- FERBER, R. and W. Z. HIRSCH (1982) *Social Experiments and Social Policy*. Cambridge: Cambridge Univ. Press.
- FEINBERG, S. E., B. SINGER, and J. M. TANUR (1984) "Large scale social experimentation in the U.S.A.," Working paper, Department of Statistics, Carnegie-Mellon University.
- HOLLAND, P. W. and D. B. RUBIN (1983) "On Lord's paradox," in H. Wainer and S. Messick (eds.) *Principles of Modern Psychological Measurement: Festschrift for Frederic M. Lord*. Hillsdale, NJ: Lawrence Erlbaum.
- KASSBAUM, G., D. WARD, and D. WILNER (1971) *Prison Treatment and Parole Survival*. New York: John Wiley.
- LENIHAN, K. (1977) *Unlocking the Second Gate*. Washington, DC: Government Printing Office.
- MOSTELLER, F. and M. C. WEINSTEIN (1984) "Toward evaluating the cost-effectiveness of medical and social experiments," in J. A. Hausman and D. A. Wise (eds.) *Social Experimentation*. Chicago: University of Chicago Press.

- PRATT, J. W. and R. SCHLAIFER (1984) "On the nature and discovery of structure." *J. of the Amer. Statistical Assn* 79: 9.
- RAUMA, D. and R. A. BERK (1982) "Crime and poverty in California: some quasi-experimental evidence." *Social Sci. Research* 11: 318.
- RIECKEN, H. W. and R. F. BORUCH (1974) *Social Experimentation: A Method for Planning and Evaluating Social Intervention*. New York: Academic.
- ROSENBAUM, P. R., and D. B. RUBIN (1984) "Comment on the nature and discovery of structure." *J. of the Amer. Statistical Assn* 79: 26.
- (1983) "The central role of propensity score in observational studies of causal effects." *Biometrics* 70: 41.
- ROSSI, P. H. (1982) "High School achievement: Public, Catholic, and private schools compared," *Amer. J. of Education* 91: 78-89.
- ROSSI, P. H. and K. LYALL (1976) *Reforming Public Welfare*. New York: Russell Sage.
- ROSSI, P. H. and H. E. FREEMAN (1982) *Evaluation: A Systematic Approach*. Beverly Hills, CA: Sage.
- ROSSI, P. H., R. A. BERK, and K. J. LENIHAN (1980) *Money, Work and Crime: Experimental Evidence*. New York: Academic.
- SHERMAN, L. W. and R. A. BERK (1984) "The specific deterrent effects of arrest for domestic assault." *Amer. Soc. Rev.* 49: 261.
- SPENCER, B. D. (1980) *Benefit-Cost Analysis of Data used to Allocate Funds*. New York: Springer-Verlag.
- TAGGART, R. (1972) *The Prison and Unemployment: Manpower Programs for Offenders*. Baltimore MD: Johns Hopkins Press.
- TANUR, J. M. (1983) "Methods for large-scale surveys and experiments," in S. Leinhardt (ed.) *Sociological Methodology 1983-1984*. San Francisco: Jossey-Bass.
- ZEISEL, H. (1982) "Disagreement over the evaluation of a controlled experiment." *Amer. J. of Sociology* 88: 378.