

THE
OXFORD
HANDBOOKS
OF
POLITICAL
SCIENCE

GENERAL EDITOR
ROBERT E. GOODIN

EDITED BY
MICHAEL
MORAN
MARTIN
REIN
ROBERT E.
GOODIN

≡ The Oxford Handbook of
PUBLIC POLICY

Since the late 1960s, spending on trials of social policy proposals in the USA has consumed over a billion dollars (Burtless 1995). In this chapter we consider the nature of social experiments that have been conducted in the past forty years. We review the efforts of many social scientists and economists to develop systematic empirical evidence about the likely advantages and disadvantages of specific policy proposals through the conduct of social experiments. Then we examine the advantages and disadvantages of social experiments themselves and try to project the current trend line into the hazy future.

2. DEFINITION

Social experiments are randomized field trials of a social intervention. Within that rubric, two emphases jostle for primacy (and a third emphasis tags along). Some authors define social experiments (SE) by emphasizing the “trial” in randomized field trial. For them, the hallmark is that a prospective intervention is being tried out on a small scale before it is widely adopted. Not only is it being tried out; it is being studied in its pilot version. The aim is to find out whether the intervention achieves its aims. If so, the assumption is that policy makers should adopt it on a system-wide basis. There is a sense of self-conscious intention to influence policy, and often this intention is accompanied by a sense of urgency as the policy window opens.

Other authors put the stress on randomization. It is randomization that allows experimenters to have confidence that the intervention was the *cause* of whatever changes are observed. In a randomized study, the experimenters select samples from the same population, assign one to the intervention, or “experimental” condition, and the other to a “control” condition. At the end of the period, the groups are compared. Inasmuch as they were very much the same at the start and the only thing that differed over time was exposure to the intervention, any differences at the end are due to the intervention. From a methodological point of view, randomization gives experimenters confidence in their estimates of effects.

The third focus in the definition of social experiments, now widely taken for granted, is that the trial is done in the “field.” Gone is the comfortable milieu of the laboratory for studying outcomes. Rather the social scientist conducts the studies in the precincts in which the actual policy will be run. Thus we have randomized field trials.

If the emphasis on randomization is accepted as the guiding principle, then any study of desired outcomes conducted through randomization is an SE. Such a definition sweeps in large numbers of evaluations of existing programs. Many evaluations of social programs are conducted after the programs are enacted, and some of the evaluations (although not nearly as many as evaluators would like) randomize prospective participants into “experimental” and “control” groups. After a period of time, the evaluator compares the status of the two groups on the desired

indicators (e.g. health status, earnings, school graduation). To blanket such *post hoc* evaluations into the category of SEs widens the category substantially.

If we confine ourselves to randomized studies undertaken on a test basis to guide adoption of future policy, we have a more focused field of enquiry. It is the definition we adopt here. Of course, the distinctions are not hard and fast. Some evaluations of existing programs are expected to guide future iterations of the program—i.e. to lead to modifications and improvements in the intervention. Sometimes, as in cases where a program at the state level is a possible model for federal policy (states as “laboratories of democracy”), what is an evaluation at one level is an SE at another. Still, the distinction is useful to hold on to. It is important to consider the main purpose for which the SE is done as well as its research design.

3. HISTORY

With a little difficulty we could probably trace SEs back to Francis Bacon, but it is sufficiently historical to go back to Sidney and Beatrice Webb. In their 1932 book, *Methods of Social Study* they argue for scientifically based social policy in words that have remarkable resonance for our own times. They advocated research conducted by social scientists trained in experimental methods who conduct independent social investigations and transmit their results to those making social policy. The actual methods, as Ann Oakley (1998a) has pointed out, were developed by educationalists and psychologists in the USA in the late nineteenth and early twentieth centuries. The philosopher Charles S. Peirce, the father of “pragmatism,” introduced the idea of randomization into psychological experiments in the 1880s. Some of the early studies dealt with the transferability of memory skills from one subject to another (Oakley cites Thorndike and Woodworth 1901 and Winch 1908). These psychological researchers invented techniques for randomly assigning subjects to experimental treatments. R. A. Fisher who did his research in agriculture and developed much that has become commonplace in statistics, is widely known for championing randomization methods.

With regard to the “field” aspect of policy experiments, Oakley (1998b) reminds us that two US sociologists, Stuart Chapin at the University of Minnesota and Ernest Greenwood at Columbia University, applied experimental methods to the study of social problems in the early years of the twentieth century. Where psychologists tended to work in laboratory settings, pioneering sociologists took their research out into the community. Chapin (1947) describes nine experimental studies that he and others carried out on topics such as recreation programs for delinquent boys, social effects of public housing, and effects of student participation in extracurricular activities. Where others had stated that randomized experiments could be done only under antiseptic laboratory conditions, he was interested in demonstrating

that they could be adapted to community settings as well. Greenwood provided a theoretical rationale for applying experimental methods to social issues, described in his book *Experimental Sociology* (1945).

In the first half of the twentieth century, most of the forerunners of current SEs were evaluations of existing programs. They shared many of the characteristics of experiments, but dealt with programs that were already up and running. The intent, nevertheless, was very similar: to see whether a program worked and, if it proved successful to extend and expand it. One evaluation that gained a great deal of attention was the Perry Preschool Project, largely because the preschool participants were followed up into their late twenties and because their lives turned out to be significantly more successful than the lives of kids in the control group (Schweinhart, Barnes, and Weikart 1993). The data provided much of the justification for authorization and reauthorizations of the Head Start program and other early childhood programs. Among other noteworthy early studies were the Eight Year Study of progressive high schools, conducted by Ralph Tyler (unpublished), the Cambridge–Somerville youth worker program that aimed to prevent juvenile delinquency (Powers and Witmer 1951), and the Hawthorne studies of reforms to working conditions in a Western Electric plant (Roethlisberger and Dickson 1939).

A relatively small number of evaluation studies used randomization for assigning participants, but some of them sought to introduce controls in other ways. Campbell and Stanley (1966) wrote a landmark monograph, *Experimental and Quasi-Experimental Designs for Research*, classifying the designs of studies that had been reported. In the language of the time, “experimental” meant that the study had randomly assigned participants to the program (or several variants of the program) and to a control group that did not receive the program. “Quasi-experimental” designs used other strategies to reduce the threat that something other than the program was the *cause* of whatever differences appeared between the groups. Although perhaps not its intent, the Campbell and Stanley book tended to legitimize quasi-experiments for evaluation purposes. Campbell and his collaborators in subsequent versions of the book (Cook and Campbell 1979; Shadish, Cook, and Campbell 2002) have sought to overcome the impression and place randomization back in priority position.

It wasn't until after the Second World War that the three main ideas of SE were combined in large-scale investigations—randomization, study in the field, and intentional preparation for policy change. With the War on Poverty in the 1960s, SEs began their modern history. The first noteworthy SE of the period was the series of income maintenance experiments. They began in 1968 in four sites in New Jersey and were followed by parallel studies in a series of urban and rural locations. The program was an effort to change the existing welfare system by the provision of a guaranteed annual income to poor people (Cain and Watts 1973; Kershaw and Fair 1976; Danziger, Haveman, and Plotnick 1981). The aim of the *experiment* was to test a policy innovation *prior* to enactment.

The income maintenance experiment was followed by experiments with housing allowances (Carlson and Heinberg 1978; Friedman and Weinberg 1983; Kennedy 1980), health insurance (Newhouse 1993), performance contracting in education

(Rivlin and Timpane 1975), and job search (Wolffhagen 1983). Greenberg and Shroder (1997) provide reports on 143 SEs conducted in the USA, one in Canada, and one in the Netherlands. All of them were randomized field trials of prospective new policies (although the policies studied in the later experiments generally represented merely incremental changes in existing programs). Only experiments that had reported results by 1996 are included in the inventory. Their appendix lists seventy-five SEs then still in progress.¹

To ground the reader in some real examples, Table 39.1 provides information on four SEs which we refer to in the following discussion.

Income maintenance experiments. Four income maintenance experiments were run in the 1960s and 1970s at eleven sites to test the impact of variations in a negative income tax program for low-income families. Families were provided with a guaranteed level of benefits and were allowed to earn additional income through work. Program benefits were reduced by a set fraction for each dollar earned. The findings showed that families reduced the number of hours they worked but not by significant amounts. Other results were mixed, with small positive results on many measures. However, by the time results were reported, the political climate had changed. Congress was in no mood to give the poor a blank check. The long and hugely expensive experiment (Greenberg and Shroder 1997 report the cost as \$111.7 million) had little policy impact.

The health insurance experiment conducted by the RAND Corporation tested the effects of varying levels of cost sharing on the use of health services and health outcomes. It randomly assigned families to one of fourteen fee for service plans or an HMO. A total of 7,708 individuals were tracked in six sites chosen to represent the United States over a period of eight years, making the experiment one of the largest and most expensive in American history. The findings showed that overall, cost sharing reduces use of medical services without substantial negative effects on health. This proved to be a factor in later acceptance of cost sharing as a cost containment strategy in both public programs and private insurance plans.

Welfare-to-work programs. In the 1980s, the Manpower Development Research Corporation (MDRC) tested ten specific state programs using random assignment, measuring the impacts and benefit–costs of state welfare-to-work programs, as well as studying their implementation. State and local governments designed, implemented, and operated the programs that were evaluated, and the MDRC developed the evaluation design and conducted the actual evaluation. The findings showed that the tested programs increased earnings and reduced the size of the welfare rolls, the benefits to society as a whole exceeded the social costs of the programs, and the programs usually resulted in net savings for taxpayers. However, the effects were relatively small.

Nursing home incentive reimbursement experiment. This experiment, conducted from 1980 to 1983, tested the effects of incentive payments for proprietary nursing homes.

¹ A new updated edition of the inventory of social experiments was published in 2004, after we had finished this chapter.

Table 39.1 Four selected social experiments

| Experiment | Tested intervention | Design | Results | Dissertation |
|--|---|--|---|---|
| Income maintenance experiments (1968–78) | Income supplements for welfare recipients with varied tax rate for paid work | Randomly assigned families to varying benefit reduction rates in 11 states | Payment of income subsidies slightly reduced number of hours worked | Welfare published in books, journal articles, and reports |
| RAND health insurance experiment (1974–82) | Varied cost sharing for medical services | Randomly assigned families to different cost-sharing programs in 6 states | Increases in cost sharing reduced use of health services without significant effect on health status | Numerous published dissertations |
| MDRC welfare-to-work experiments (1975–88) | Provided job training and other employment services to AFDC participants | Randomly assigned AFDC recipients to various employment programs in 10 states | Consistent small positive effects on participants' earnings, reductions in welfare rolls and in cost to taxpayers | Welfare dissertations during welfare debates |
| Nursing home incentive reimbursement experiment (1980–3) | Provided reimbursement incentives for nursing homes accepting Medicaid patients | Randomly assigned 36 nursing homes to participate in intervention program or control group | Little effect of reimbursement on health outcomes or discharge of Medicaid patients. Slight increase in admissions of heavy care patients | Not widely dissertated |

The aim was to encourage them to accept more hard-to-care-for Medicaid patients and to discharge patients to lower-care facilities when they had attained acceptable health status. The study was conducted with a total of thirty-six nursing homes in San Diego County, eighteen of which were in the control group. Findings showed that in the first year of the experiment there was no difference between the two groups of nursing homes in the intensity of care that admitted patients required, but in the second year the experimental nursing homes did admit patients in need of more intensive care. No statistically significant differences emerged on achievement of patient health goals or on patient discharges to less expensive facilities. The small size of the sample and the shortness of time over which the experiment was run (thirty months) militated against significant differences. The findings were not disseminated widely, and few people heard about the results.

4. THEMES

It seems obvious that social experiments (SEs) are conducted to improve decision making regarding policies under study. However, a direct relationship between the results of SEs and policy decisions presumes a rational policy environment with established pathways for information from experiments to feed into policy decisions. The relationship between the conduct of SEs and the policy environment is more complex than such a simple statement suggests. SEs are generally lengthy and results arrive in changed, sometimes unreceptive policy space. Experiments arise for a variety of reasons and are not always set up to answer directly specific policy questions. And indeed experiments are but one in a multitude of information sources that policy makers must consider when making policy decisions.

In this chapter, we explore the relationship of SEs to policy making. First we look at the advantages of conducting such experiments. We examine contributions to policy and contributions to social science. Then we describe the disadvantages that SEs entail both for the policy process and for social science. Last, we puzzle about their future, in a near-sighted attempt to foresee what use is likely to be made of SEs as political and economic conditions change.

We admit that our view is largely a United States view, but that is not totally our doing. The story of SEs has been largely a US story. The first large experiments were done in the USA and most of the subsequent work has been “made in the USA.” In recent years, Canada has jumped on the bandwagon, and the Netherlands has also conducted a few experiments. But most of the experience on which the policy world relies is US work.

Running alongside our discussion of advantages and disadvantages of SEs are three main themes. Hold the pages sidewise and you will see these ideas: (1) The policy world is a complex place. Policy making evolves from ideologies and beliefs, interests,

and institutional norms, as well as from competing information. “Scientific evidence” alone will almost never determine the direction of policy making. (2) The research world is no less complex. Technical issues bedevil the study of complex policy issues and affect the extent to which social scientists can derive authoritative evidence. (3) The fit between the worlds of policy and research is inexact. Sometimes the answers that SEs provide bear little resemblance to the questions that decision makers ask. A major misalignment is timing. An experiment may not be completed until long after the questions that provoked the experiment have faded from view. Another issue is the uneasy pattern of communication between researchers and policy makers. Nevertheless, despite all the disabilities that affect SEs, we conclude that a well-done SE provides important information that illuminates the policy field and has at least the potential for influencing policy.

5. ADVANTAGES OF SOCIAL EXPERIMENTS

5.1 Policy Advantages

Provide Data on Likely Outcomes of a Policy Idea

Social experiments are experimental tests of new policy ideas. They provide information to people engaged in the political process of making policy. They advance the *rational component* in policy making (Rivlin 1971). Many policy decisions are made in a relative information vacuum with little known about the actual effects of the policies proposed. Data from well-designed tests of policies under discussion can provide invaluable information about the realities of the expected effects of policy adoption, including the potential for unexpected or negative consequences. In some cases, such information has counted in decisions to adopt a particular policy track. For example, the positive results of the welfare-to-work experiments played a modest role in the further expansion of work requirements in state welfare programs. In addition, the success of state-designed and -implemented welfare-to-work programs may have encouraged later legislation to give states flexibility to design state-specific welfare programs (Greenberg, Linksz, and Mandell 2003; Baum 1991).

Some advocates claim that SEs offer objective information, unsullied by the pull of interests. But objectivity is relative. Social scientists for over a generation have acknowledged that every social science enquiry is inevitably colored by the assumptions, biases, and blinkers of its investigator. Nevertheless, experiments appear less prone to dispute than most other forms of knowledge. They collect information systematically from a known population according to the canons of social science. The element of randomization adds authoritativeness. When there is contention, other social scientists can reanalyze the data to try to support their argument. In resolving disputes, SEs rely on the judgement of the community of social scientists.

(See Howell and Peterson 2004; Krueger and Zhu 2004, on rival interpretations of school choice experiments.) On any reasonable scale, experimental information is credible. In the four experiments that we have cited here, little important disagreement emerged about the interpretation of the findings.

Clarify Trade-offs

Social experiments can at times clarify the key trade-offs in policy decisions and provide information to debate these trade-offs (Orr 1998). For example, the AFDC Homemaker Home Health Aide Demonstration found that home care did not reduce health costs but did improve clients' sense of well-being. The findings provided policy makers with information to debate the trade-off between the costs and benefits of the program.

Keep a Policy Idea Alive

One aim ascribed to social experiments is keeping alive a policy idea that cannot muster enough support at the moment to ensure passage. The income maintenance experiments were reportedly undertaken because most members of Congress did not support a negative income tax for the poor to replace the welfare system. The federal Office of Economic Opportunity and academic economists who favored the idea could not carry the day, but they gained support for an experiment (and then additional experiments) in the hopes of making a good case. They might also have hoped that the political winds would change, and members of Congress would come to embrace their idea for income maintenance for the poor. (Despite their efforts, the negative income tax was not to be.)

The contrary assumption, that SEs are used to delay a new policy until the lengthy study is done, does not receive much empirical support. Once a policy proposal has acquired political momentum, it is usually enacted regardless of evidence. Before results were available from the housing allowance experiments, Congress enacted one feature that was still being tested. They passed a bill, known as Section 8, that provided subsidized payments for the poor in the private housing market.

Stock a Library of Information

SEs can create inventories of information for future policy situations (Feldman 1989). Although their sponsors, with their eyes focused on current options, do not intend only to pile up knowledge for the future, that is one likely result. Even if the findings of the experiment have little impact on current discussions, they do provide a stock of information that future political actors and analysts can draw on (Orr 1998). For example, the health insurance experiment notably provided information on elasticities in health care demand that informed later analysis.

Help to Build Consensus

The focus and intensity of a social experiment, coupled with a general acceptance among researchers of the quality of impact estimates derived through experimental designs, may provide the focal point needed to draw together diverse actors and information sources to agreement. The health insurance experiment finding that cost sharing reduced health care use without harming health led to a fairly broad acceptance among researchers and policy makers of cost sharing as a legitimate cost containment strategy. Similarly, the welfare-to-work experiments broadened acceptance of mandated work requirements in public assistance programs.

Legitimize Existing Preferences

If the results of an experiment align with preferences of decision makers, they can provide legitimacy to existing policies or preferred alternatives. They can reaffirm policies after the policy has been chosen (Greenberg and Mandell 1991). Some social scientists worry that this kind of after-the-fact legitimization is a misuse of social science. But if the findings support a policy that policy actors have already selected on other grounds, there doesn't seem anything wrong with giving it a social science seal of approval.

At times, social experiments may provide political cover for either difficult or highly contested policy decisions, shifting the onus of decision making onto "science." They may offer policy makers a set of data-driven arguments for or against a particular policy option.

5.2 Research Advantages

Spur the Development of New Research Methods

In order to do the challenging work of SEs, social scientists have had to develop new methods and techniques. They have also had to develop new statistical methods to analyze the data. The field environment, the size of the samples, the rarity of certain groups about whom data is needed, the need to generalize to a larger population, the need to measure difficult concepts—all have contributed to innovations in research methods. Current textbooks bear witness to the methodological advances spurred by decades of social experimentation.

Real-life Test for Social Theories

Another advantage for social science is that SE gives social scientists the opportunity to test theories in the crucible of real-world settings. They can subject theories and practices based on those theories to actual test. This can help bring abstract theorizing down to a practical level. For example, theories about the value of competition

in improving the quality of schools are being tested in a number of SEs that give parents choice of their children's schools (Howell and Peterson 2004). Theories about the positive effects of a non-stigmatizing guaranteed income, implemented through a negative income tax, were studied in urban and rural areas for extensive periods of time.²

Many of the pilot ideas that SEs have studied originated not in social science theories but in political or practice settings. For example, the MDRC welfare experiments did not directly test any specific behavioral theory. Nevertheless, they often derived from—or coincided with—theories that were current among social scientists. The studies therefore supported, refuted, or failed to provide convincing evidence regarding the theories to which they were related.

Provide Interesting Work to Social Scientists

SEs are interesting, frontier studies. They generate considerable enthusiasm among social scientists, especially those who work in research institutes that have the resources to do them well. SEs require skilled staff and the latest statistical know-how to do this kind of demanding work, and only a few organizations have over time been able to establish and maintain the type of expertise needed for such work. An analysis of the 143 SEs identified in *The Digest of Social Experiments* found that three organizations dominate the conduct of SEs in the USA: Abt Associates, the Manpower Demonstration Research Center (MDRC), and Mathematica Policy Research conducted almost half of the experiments reviewed (Greenberg et al. 1999). In Canada, the Social Research and Demonstration Association does most of the social experiments.

One of the interesting things about SEs is that economists are the investigators in most of them. Economists, who haven't been known for their empirical fieldwork, in a sense reinvented survey research for the income maintenance experiments, and developed sampling and analysis techniques from their tradition. Why economists? Many of the topics deal with money. They are testing schemes that expect to reduce government expenditures. Do welfare-to-work programs reduce the welfare rolls and welfare costs? Does nursing home reimbursement increase intake of patients in need of intensive care so that they do not have to stay in (very expensive) hospitals? Do job-finding programs reduce the length of time that unemployed workers receive unemployment compensation? Another reason for the frequent presence of economists is that money is easier to measure than the outcomes that often concern sociologists and psychologists, such as "functional ability" or "age-appropriate childhood development." Policy makers and the public find data on costs and savings more credible than fuzzier concepts. Economists have the techniques to study and model data denominated in dollars.

² See Kershaw and Fair 1976; Watts and Rees 1976; Palmer and Peckman 1978.

6. LIMITATIONS OF SEs

6.1 Policy Limitations

Effects on Decisions

When we review the history of social experiments, we see that they have not had a decisive, direct effect on the ensuing decisions. Of our four examples, only the welfare-to-work experiments were later reflected in policy. Neither the health insurance experiment, the nursing home incentive reimbursement experiment, nor the income maintenance experiments made much of a dent at all, and the findings were relegated to the great analytical storehouse. Even in the welfare-to-work experiments, where experiment results seemed to affect later policy, the result was at best indirect.

Greenberg, Mandell, and their colleagues did a telephone interview study of welfare directors in the states. They found that while most of the state directors knew something about the findings of the welfare-to-work experiments (although not the specifics), they didn't believe the findings had influenced the policies of their own state. What they did value was the demonstration that states could administer the program without much problem and a general sense that work first was better than training first for former welfare recipients. In their 2003 book, Greenberg et al. conclude:

Ironically, however, even though these experiments did have important effects on policy, their role was nonetheless limited . . . In particular, many policymakers already viewed the programs tested by the welfare to work experiments as attractive on other grounds. Findings from the experiments simply reinforced that view. Consequently, rather than being pivotal to whether the types of programs they tested were adopted, they were instead used persuasively and in designing these programs. In other words, they aided policymakers in doing what they already wanted to do. (2003, 308, 310)

Why should the results of SEs be so marginal? Why doesn't rationality reign?

Social scientists are under no illusions that "scientific evidence" will displace all other sources of understanding. Policy making is also based on ideologies and beliefs, interests, competing information, and institutional norms (Weiss 1983, 1995). The results of social experiments can nudge policy only a small distance, and their influence is dependent in large part on the interplay with the other factors in the policy environment. Social scientists know that legislators and administrative officials have long-standing beliefs and principles that guide much of their orientation toward policy. Their ideological orientation exerts powerful influence over which policy proposals receive even a hearing. Attitudes toward abortion and gay marriage are obviously determined by ideology and principles, but it is not only on such extreme issues that ideology often prevails. For some policy makers, similarly strong beliefs affect their views of the enactment of a draft, the need for standardized performance tests in schools, mandatory sentences for repeat offenders, and needle exchange programs for drug addicts.