

- Tufte, Edward R. 1974. *Data Analysis for Politics and Policy*. Englewood Cliffs, NJ: Prentice-Hall.
- Tufte, Edward R. 1983. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.

## **The Experimenting Society**

*Donald T. Campbell*

### **Some Ideology of Science**

Science requires a disputatious community of “truth seekers.” The ideology of the scientific revolution agrees with Popper’s (1952, 216–22) epistemological sociology of science. The norms of science are explicitly anti-authoritarian, antitraditional, antirevelational, and pro-individualistic. Truth is yet to be discovered. Old beliefs are to be systematically doubted until they have been reconfirmed by the methods of the new science. Persuasion is to be limited to equalitarian means, potentially accessible to all: visual demonstrations and logical demonstrations. The community of scientists is to stay together in focused disputation, attending to each others’ arguments and illustrations, mutually monitoring and “keeping each other honest,” until some working consensus emerges (but mutual conformity in belief per se is rejected as an acceptable goal). Note how the ideology explicitly rejects the normal social tendency to split up into like-minded groups on specific scientific beliefs, but at the same time it requires a like-mindedness on the social norms of the shared inquiry. Sociologically, this is a difficult ideology to put into practice. Merton (1973) has phrased the requirement as “organized skepticism.” Both features are required, yet “organized” and “skepticism” are inherently at odds. Societal and institutional settings in which it can be approximated are rare and unstable. Nonetheless, it may be regarded as a viable socio-

logical thesis about a system of belief change that might plausibly improve beliefs about the physical world, including especially the not-directly-observable physical world.

Such a social system could not have emerged in just any problem area. It required collective success experiences, built around novel theory related to visually compelling demonstrations that could be independently replicated time and time again. It no doubt helped that many of the problems (as in static and magnetic electricity) were ones upon which political and religious powers held no important fixed beliefs. The social sciences are not an arena in which the social system of successful science could have first emerged. If pure or applied social studies are to merit the term scientific, their problem areas will have to be “colonized” from the successful sciences. Such colonization will be dependent upon a valid theory of the social system of validity enhancing belief change of the successful sciences (Ravetz 1971; Campbell 1986a).

### The Applied Social Science of Program Evaluation

My central concerns over the past twenty-five years (my concerns, at least, as known to my fellow social scientists) have been as a methodologist trying to extend the epistemology of the experimental method into nonlaboratory social science. They know me for my lists of threats to validity in quasi-experimental research and for my list of design alternatives that will help render these threats less plausible. (Campbell 1957a, 1979a; Campbell and Stanley 1963; Cook and Campbell 1979). Since 1969 (Campbell 1969e) at least, this concern has been focused on applied social science, on treating the ameliorative efforts of government as field experiments. In the American social science community, this falls within “policy research.” My aspect of it is known as “program evaluation.” I am reporting here on our experience in developing methods for program evaluation and their implications for the question, “Can the open society be an experimenting society?”

Two aspects of program evaluation research push us toward speculating about an experimenting society. On the one hand, as we try to implement high-quality program evaluations, we meet with continual frustration from the existing political system. It seems at times set up just so as to prevent social reality-testing. This leads us to think about alternative political systems better designed for evaluating new programs.

On the other hand, if we look at our own recommendations to government regarding how to implement programs so that their impact can be evaluated, one can see that we evaluation methodologists are, in fact, often proposing novel procedures for political decision making. We are, in fact, designing alternative political systems. If we were self-consciously aware of this, we would, I believe, often make different recommendations.

Out of these influences comes the imagery of an experimenting society, one that would vigorously try out possible solutions to recurrent problems and would make hard-headed, multidimensional evaluations of outcomes, and when the evaluation of one reform showed it to have been ineffective or harmful, would move on to try other alternatives. There is no such society anywhere today. While all nations are engaged in trying out innovative reforms, none of them are yet organized to adequately evaluate the outcomes of these innovations.

### The Experimenting Society and Utopian Thought

While benefiting, I hope, from the twenty years of self-conscious American discussion and practice, this is, nonetheless, a speculative exercise in utopian thought. We should bear in mind, as we consider it, all of the problems of utopian thought that Marx (even though in the end he exemplified these problems rather than avoided them) and Popper have noted. Several features of this model seek to avoid those problems—how well, we must discuss.

First, a critical utopianism is intended. While I call upon a number of you to make a career commitment to being theorists and methodologists of the experimenting society, I charge you to do this and not only to be ready with appropriate scientific methods for evaluating new programs when the political will for an experimenting society appears, *but also to contribute to the best possible exploration—in advance of what such a society would be like*, with the moral commitment being made right now that this most thorough exploration *may lead you and I in the end to oppose the experimenting society, and that this outcome too would thoroughly justify our career commitment*, as a morally responsible and socially necessary function. My own explorations have raised many problems which (even at the a priori utopian level) we have not solved. Until we have more plausible solutions, I myself withhold full advocacy.

This is, I argue, very much in Popper's spirit of "letting our ideas die instead of ourselves," conjectures and vigorous mutual efforts at refuting, selecting, revising, and purifying our ideas in advance of trying them out, particularly where those practical trials are bound to be costly and socially disruptive in ways that might in part have been anticipated.

While my own work is focused on this one utopia (the experimenting society), I recommend that it be seen as part of that important perspective of modern future studies, namely the strategy of alternative futures. We are to elaborate and conceptually explore, with mutual criticism, a variety of alternative utopias, choosing—tentatively to guide our political action—the one that compares best with the others in as complex and wise a comparison as our detailed elaboration and mutual criticism can now provide. (How tragic that the journal *Alternative Futures* died after only five years of publication. Fortunately, we still have other "futures studies" journals.) In the alternative futures strategy for guiding current political thought, validity can not be achieved in the form of true theories established at one time as absolutely true and superior to all potential alternatives. Rather, we must live by theories fallibly selected from among the few alternatives explicitly available in the contemporary competition among alternatives. Let us make this choice as competently as possible through the most thorough, critical conceptual exploration of the widest range of alternatives.

Second, the experimenting society is a process utopia, not a utopian social structure per se (Schwarzlander 1978; Haworth 1960). It seeks to implement that recommendation of Popper's, "A social technology is needed whose results can be tested by piecemeal social engineering" ([1945] 1952, v. 2, 222). The implication of this feature in avoiding the errors of most utopian thought will be explored in more detail in what follows.

#### *The Ideology of the Experimenting Society*

Before getting into the methodological details, a little more needs to be said about the nature of the experimenting society as an ideology for a political system.

It will be an *active society* (Etzioni 1968) preferring exploratory innovation to inaction. It would be a society which experiments, tries

things out, explores possibilities in action (as well as, or even instead of, in thought and simulation). It would borrow from epistemology and the history of science the truism that one cannot know for certain in advance, that a certain amount of trial-and-error is essential. Faced with a choice between innovating a new program or commissioning a thorough study of the problem as a prelude to action, the bias would be toward innovating. It will be committed to action research, to action as research rather than research as a postponement of action (Lewin 1946, 1948; Sanford 1970). It will be an evolutionary, learning society (Dunn 1971).

It will be an honest society, committed to reality-testing, to self-criticism, to avoiding self-deception. It will say it like it is, face up to the facts, be undefensive and open in self-presentation. Gone will be the institutionalized bureaucratic tendency to present only a favorable picture in government reports. For many a civil servant on both sides of the Iron Curtain this freedom to be honest will be one of the strongest attractions of the experimenting society. The motive of honesty in political reform, revolution, and personal heroism has been generally neglected until recently. It is of course now a dominant theme among our young idealists, showing up in their standards for their own interpersonal relations and in their criticism of the cowardly hypocrisy, double-talk, and dishonesty of their elders. It also emerges as a major political force within the communist countries, as Polanyi (1966a) has argued in the case of the Hungarian uprisings of 1956. While that revolution no doubt had complex roots, many of which might accurately be called reactionary or fascist, Polanyi persuades me in his analysis of the motives of the "Petofi Circle." These elite communist journalists, well rewarded by their establishment with power and wealth, were motivated by the pain of continually having to write lies and by the promise of a society in which they could write the truth as they saw it.

Honesty was also among the prime motives of the Czechoslovak leadership and followership of 1968, accounting for both its great popularity and its dysfunctionally provocative excesses in exposing past lies. This sounds more unsympathetic than I feel. While I share the belief that more moderation in press statements, in the program of retrials, and in other corrections of past falsehoods probably would have avoided the external interventions while allowing the continuation of the more tangible innovations, I also sympathize with the spirit

that rejected such compromise. In October 1968, I, along with fifty others, attended the International Conference on Social Psychology at Prague. The numerous Czechoslovak people we met, while profoundly distressed by the occupation, were still glowing with the excitement of their January to August experiment and were eager to talk. They freely conceded that any economic reforms that might have been initiated could not have had any effect in that short time and perhaps would not have even in the long run. But the honesty reforms had immediate effects and were profoundly enjoyed. These were most obvious, perhaps, in newspapers, television, and radio, which became exciting as specific reporters were really allowed to describe things to the best of their personal beliefs and knowledge. It is hard to exaggerate the contrast this represented over the prior times in which carefully worded, multiply vetoed statements said only what was wanted said and even disguised policy changes under ritual jargon. As it affected personal lives, the honesty reforms took the shape of dismantling the thought-police apparatus, cessation of persecutions because of beliefs of statements, reestablishment of falsely maligned reputations, and full exposure of past lies of the state. Honesty was a great part of Dubcek's amazing personal popularity. Here was a weak compromise candidate whose past history gave no more promise of greatness than did Harry Truman's in the United States, for example. He had been great in how he carried out a role, rather than being one who obtained the role through greatness. Prior to Dubcek, every official appearing on television read carefully from a multiply censored, cautiously expressed, and often dishonest text. Dubcek instead spoke freely without notes, describing things as he saw them, naively expressing honest emotions in words, facial expressions, and tone of voice.

As one came to know how decisively (and flagrantly) they had been challenging, changing, and criticizing the past regime, and exposing its cruelty and dishonesty, one naturally wondered if they would not have been wiser to have moved more slowly and less provocatively and thus have avoided the occupation and retained moderate gains. But this was our question as outsiders. The issue was never raised by them in anything like these terms, even though they were uniformly deeply pessimistic over what was in store for them. When one raised the question, they said: "Perhaps so, but Dubcek had no choice. He was the most conservative and moderate of the democratic majority." The real answer, one felt, was that their joy and pride in their brief,

outspoken, free, and optimistic period made it something they would not want to have missed, that it had a great value that compromise would have spoiled, that it was an experience worth the price of the risk they ran, even when viewed after the gamble had been lost.

Their reform as they saw it was within communism or socialism, not at all changing the ownership of the means of production, and quite interpretable as compatible with Marx and Lenin. (A quotation from Marx was widely displayed which damned large states for using their power to impose their will on small states.) Their slogan was, "Communism with a human face," a humane communism, a democratic communism. Their main target was inhumane bureaucracy. Their reforms were primarily in the location of authority: workers' councils for factories, local authorities as decision-makers and censors, etc. Coming as it did from within the party apparatus as a majority position, and inspired by their most gifted Marxist writers, the movement served to create a new enthusiasm for worn-out ideology. A revitalized exportable evangelical communism might well have resulted, appealing to the disenchanted in both capitalist and communist countries.

The experimenting society will be a nondogmatic society. While it will state ideal goals and propose wise methods for reaching them, it will not dogmatically defend the value and truth of these goals and methods against disconfirming evidence or criticism.

It will be a *scientific society* in the fullest sense of the word "scientific." The scientific values of honesty, open criticism, experimentation, willingness to change once-advocated theories in the face of experimental and other evidence will be exemplified. This usage should be distinguished from an earlier use of the term "scientific" in social planning. In this older usage, one scientific theory of society is judged to be established as true. On the basis of this scientific theory, extrapolations are made to the design of an optimal social organization. This program is then put into effect, but without explicit mechanisms for testing the validity of the theory through the results of implementing it. Such social planning becomes dogmatic, nonexperimental, and is not scientific in the sense used here even though the grounds of its dogma are the product of previous science. Such dogmatism has been an obvious danger in implementations of Marxist socialism. It is also a common bias on the part of governmental and industrial planners everywhere. In such planning, there is detailed use of available science but no use of the implemented program as a check on the validity of

the plans or of the scientific theories upon which they were based. Thus economists, operations researchers, and mathematical decision theorists trustingly extrapolate from past science and conjecture, but in general fail to use the implemented decisions to correct or expand that knowledge.

It will be an *accountable, challengeable, due-process society*. There will be public access to the records on which social decisions are made. Recounts, audits, reanalyses, reinterpretations of results will be possible. Just as in science objectivity is achieved by the competitive criticism of independent scientists, so too the experimenting society will provide social organizational features making competitive criticism possible at the level of social experimentation. There will be sufficient separation of governmental powers so that meaningful legal suits against the government are possible. Citizens not a part of the governmental bureaucracy will have the means to communicate with their fellow citizens disagreements with official analyses and to propose alternative experiments. It will be an open society (Popper 1945).

It will be a *decentralized society* in all feasible aspects. Either through autonomy or deliberate diversification, different administrative units will try out different ameliorative innovations and will cross-validate those discoveries they borrow from others. The social-system independence will provide something of the replication and verification of successful experiments found in science. Semi-autonomy will provide some of the competitive criticism that promotes scientific objectivity.

It will be a society committed to *means-idealism* as well as *ends-idealism*. As in modern views of science, the process of experimenting and improving will be expected to continue indefinitely without reaching the asymptote of perfection. In this sense, all future periods will be mediational and transitional, rather than perfect-goal states. Ends cannot be used to justify means, for all we can look forward to are means. The means, the transitional steps, must in themselves be improvements.

It will be a *popularly responsive society*, whose goals and means are determined by collective good and popular preference. Within the limits determined by the common good, it will be a *voluntaristic society*, providing for individual participation and consent at all decision-levels possible. It will be an equalitarian society, valuing the well-being and the preferences of each individual equally.

This too-brief sketch has glossed over many problems. A few of these will be considered later on in the paper. A few need to be

considered now. While the experimenting society described thus far has many attractive features, it obviously has many dangers and costs. We methodologists for the experimenting society should be sensitive to our own and others' ambivalences about it. We should anticipate its dangers and misuses as well as its promises. We should, of course, design ways of obviating these where possible. But we should keep open the possibility that we will end up opposing it. We who now specialize in thinking about the experimenting society should be the first to decide that it is unworkable or in the net undesirable, if indeed it is and if this can be ascertained in advance of trying it.

The ideals described as characterizing the experimenting society are for the most part ideals that all of today's major ideologies claim as their own. They are endorsed in both communist and capitalist countries, most clearly in each one's criticisms of the other. While neither type of society is as yet achieving these ends, the experimenting society could grow out of either, or out of both. A competition to see which could best and soonest implement it might even be envisaged. In any event, we methodologists for the experimenting society should consider the problems of implementing it under all available forms of political-economic organization and should regard designing all such routes as a part of our methodological challenge. Such universalism is made easier both by the common ideals and by the fact that the ideology of the experimenting society is a method ideology, not a content ideology. That is, it proposes ways of testing and revising theories of optimal political-economic-social organization rather than proposing a specific political and economic system. (Of course, this is too simple, for the requirements of the experimenting society exclude many forms of political-economic-social organization, but those excluded are not being advocated by either side of the cold war.)

Once implemented in both capitalist and communist countries, one might expect that social experimentation would tend to produce increased similarity of social organization, much as industrialization has tended to do. This expectation is based upon the assumption of some universals in human preferences for the good life and the good society. Once these preferences become the selective criteria for choosing among alterations of existing forms, these universals will presumably shape societies toward a common optimum. (This needs qualifying in terms of the theory of the conditions of convergence and nonconvergence in iterative processes.)

There are, of course, forces operating against the development of the experimenting society in both capitalist and communist countries, both in ideology as well as in current practice. Thus the Marxist-Leninist commitment to the necessity of a dictatorship during the transitional phase has served in practice to justify a dogmatism and intolerance of criticism that are inconsistent with the experimenting society. Yet in the total, still-developing body of Marxist theory (perhaps particularly among Yugoslav and Polish theorists) there are stated perspectives and ideals quite sufficient to justify a truly experimental socialism.

Within Western democratic capitalism, there are a number of favorable features. These include the legal tradition, the successful achievement of changes in government through elections, and the genuine pluralism of decision-making units. The so-called "market mechanisms" of capitalist economic theory can be regarded in ideal form as self-regulatory cybernetic feedback systems implementing the collective aspects of the preferences of individual decision-makers. But the ideological justification and effective practice of the accumulation of great inequalities in individual and corporate wealth, and the role of wealth in providing grossly uneven weightings of some persons' preferences over those of others, provide great obstacles that may effectively sabotage program decisions genuinely based on the public good. Within both capitalist and communist countries there are shared aspects of political processes that work against the emergence of the experimenting society and which may, in the long run, preclude it. In the course of this paper, we will examine a number of such problems, each of which needs technical solutions before we can advocate an experimenting society.

#### *The Social Scientist as Servant of the Experimenting Society*

Societies will continue to use preponderantly unscientific political processes to decide upon ameliorative program innovations. Whether it would be good to increase the role of social science in deciding on the content of the programs tried out is not at issue here. The emphasis is, rather, on the more passive role for the social scientist as an aid in helping society decide whether or not its innovations have achieved desired goals without damaging side effects. The job of the methodologist for the experimenting society is not to say what is to be done,

but rather to say what has been done. The aspect of social science that is being applied is primarily its research methodology rather than its descriptive theory, with the goal of learning.

This emphasis seems to me to be quite different from the present role as government advisors of most economists, international relations professors, foreign area experts, political scientists, sociologists of poverty and race relations, psychologists of child development and learning, etc. Government asks what to do, and scholars answer with assurance quite out of keeping with the scientific status of their fields. In the process, the scholar-advisors too fall into the overadvocacy trap, and fail to be interested in finding out what happens when their advice is followed. Certainly the idea that one already knows precludes finding out how valid one's theories are. We social scientists could afford more of the modesty of the physical sciences, should more often say that we can't know until we've tried. For the great bulk of social science where we have no possibility of experimentally probing our theories, we should be particularly modest. While the experiments of the experimenting society will never be ideal for testing theory, they will probably be the best we have, and we should be willing to learn from them even when we have not designed them. More importantly, measuring the effects of a complex politically designed ameliorative program involves all of the problems of inference found in measuring the effects of a conceptually pure treatment variable—all and more. The scientific methods developed for the latter are needed for ameliorative program evaluation.

The distinction is overdrawn. It reflects my own judgment that in the social sciences, including economics, we are scientific by intention and effort, but not yet by achievement. We have no elegantly successful theories that predict precisely in widely different settings. Nor do we have the capacity to make definite choices among competing theories. Even if we had, the social settings of ameliorative programs involve so many complexities that the guesses of the experienced administrator and politician are apt to be on the average as wise as those of social scientists. But whatever the source of the implemented guess, we learn only by checking it out. Certainly in the experimenting society, social scientists will continue to be called upon to help design solutions to social problems, and this is as it should be. Perhaps all I am advocating in emphasizing the role of servant rather than leader is that social scientists avoid cloaking their recommendations in a spe-

cious pseudoscientific certainty and, instead, acknowledge their advice as consisting of wise conjectures that need to be tested in implementation.

The servant-leader contrast is overdrawn in other senses also. The truism that measurement itself is an agent of change is particularly applicable to the experimenting society. Advocating hard-headed evaluation of social programs is a recommendation for certain kinds of political institutions. In considering the methodological challenges of the experimenting society in what follows, appeals to the theory and content of the social sciences will be made, as well as to their methodology.

### **Methodological Problems**

#### *Resistance to Assessing the Outcome of Praxis and the Overadvocacy Trap*

In the United States one of the pervasive reasons why interpretable program evaluations are so rare is the widespread resistance of institutions and administrators to having their programs evaluated. The methodology of evaluation research should include the reasons for this resistance and ways of overcoming it.

A major source of this resistance in the United States is the identification of the administrator and the administrative unit with the program. An evaluation of a program under our political climate becomes an evaluation of the agency and its directors. In addition, the machinery for evaluating programs can be used deliberately to evaluate administrators. Combined with this, there are a number of factors that lead administrators to correctly anticipate a disappointing outcome. As Rossi (1969) has pointed out, the special programs that are the focus of evaluation interests have usually been assigned the chronically unsolvable problems—those on which the usually successful standard institutions have failed. This in itself provides a pessimistic prognosis. Furthermore, the funding is usually inadequate, both through the inevitable competition of many worthy causes for limited funds and because of a tendency on the part of our legislatures and executives to generate token or cosmetic efforts designed more to convince the public that action is being taken than to solve the problem. Even for genuinely valuable programs, the great effort required to overcome

institutional inertia in establishing any new program leads to grossly exaggerated claims. This produces the “overadvocacy trap” (Campbell 1969e), so that even good and effective programs fall short of what has been promised, which intensifies fear of evaluation.

The seriousness of these and related problems can hardly be exaggerated. As methodologists, we in the United States are called upon to participate in the political process in efforts to remedy this situation. But before we do so, we should sit back in our armchairs in our ivory towers and invent political/organizational alternatives that would avoid the problem. This task we have hardly begun, and it is one in which we may not succeed. Two minor suggestions will illustrate. I recommend that we evaluation-research methodologists should refuse to use our skills in *ad hominem* research. While the expensive machinery of social experimentation can be used to evaluate persons, it should not be. Such results are of very limited generalizability. Our skills should be reserved for the evaluation of policies and programs that can be applied in more than one setting and that any well-intentioned administrator with proper funding could adopt. We should meticulously edit our opinion surveys so that only attitudes toward program alternatives are collected and such topics as supervisory efficiency excluded. This prohibition on *ad hominem* research should also be extended to program clients. We should be evaluating not students or welfare recipients but alternative policies for dealing with their problems. It is clear that I felt such a prohibition is morally justified. But I should also confess that in our U.S. settings it is also recommended out of cowardice. Program administrators and clients have it in their power to sabotage our evaluation efforts, and they will attempt to do so if their own careers and interests are at stake. While such a policy on our part will not entirely placate administrators' fears, I do believe that if we conscientiously lived up to it, it would initiate a change toward a less self-defeating political climate.

A second recommendation is for advocates to justify new programs on the basis of the seriousness of the problem rather than on the certainty of any one answer and to combine this with an emphasis on the need to go on to other attempts at finding a solution should the first one fail (Campbell 1969c). Shaver and Staines (1971) have challenged this suggestion, arguing that if an administrator takes this attitude of scientific tentativeness, it constitutes a default of leadership. Conviction, zeal, enthusiasm, and faith are required for any effective

effort to change traditional institutional practice. To acknowledge only a tentative faith in the new program is to guarantee a half-hearted implementation of it. But the problem remains; the overadvocacy trap continues to sabotage program evaluation. Clearly, we should address our social-psychological and organizational-theoretical skills to this problem.

#### *The Use of Experimentation in the Experimenting Society*

Social scientists often argue that they, like the astronomers, can build a science on correlational evidence alone and can do without the outmoded concept of cause. For a while, major philosophers of science agreed. Recently, however, the concepts of cause and effect are being readmitted to the most sophisticated philosophy of science (see Cook and Campbell 1979, ch. 1, for a review). Whatever stand one may take on this, it must be admitted that intentional projects to improve society (*praxis*, if you will) accept both the concept of cause and the concept of learning from *praxis*. Implementing social change efforts in natural settings is a *praxis* akin to experimentation, differing only in the equivocality of interpreting the outcome. The big question of this paper is whether or not we can improve the interpretability of intentional *praxis* in our political life without destroying the open society in the process.

This is not to deny that the concept of cause may be a logical hodgepodge, involving a number of analytical criteria of no logical relationship to each other, and of no entailed status beyond observation of past correlation. Let us accept the fact that man's deeply ingrained concept of cause is a product of biology, psychology, and evolution (Campbell 1974b) rather than a pure analytic concept. If so, it reflects the adaptive advantage of being able to intervene in the world to deliberately change the relationship of objects. From among all the observable correlations in the environment, man and his predecessors focused upon those few which were, for him, manipulable correlations. From this emerged man's predilection for discovering "causes" rather than mere correlations. In laboratory science, this search is represented in the experiment, with its willful, deliberate intrusion into ongoing processes. Similarly for the ameliorative social scientist. Of all of the correlations observable in the social environment, we are interested in those few which represent manipulable relationships, in

which—by intervening and changing one variable—we can affect another. No amount of passive description of the correlations in the social environment can sort out which are "causal" in this sense. To learn about the manipulability of relationships, one must try out manipulation. The scientific, problem-solving, self-healing society must be an experimenting society. (The causal modeling of current sociology can only generate focal hypotheses still needing cross-validation in praxis [Cook and Campbell 1979, ch. 7]).

*Accepting the important role of social experimentation at once puts clarity of scientific inference into conflict with means-idealism and acceptable political processes.* Experiments involving randomized assignment to treatments are undoubtedly the most efficient and valid. While I am a strong advocate of good quasi-experimental designs where randomization is not possible, detailed consideration of specific cases again and again reinforces my belief in the scientific superiority of randomized assignment experiments. As advisors to the experimenting society we will often recommend such research designs. Yet they present special moral problems that we will have to consider. First, such experiments are best done where those designing and directing the study have most complete and arbitrary control over the people participating in the study, that is in total institutions such as prisons and armies. One needs optimally to be able to randomly assigned persons to experimental treatments and to enforce 100 percent participation in these treatments. Moreover, to avoid reactive arrangements, the participants should be unaware of the experiment, unaware that other people are deliberately being given different treatments (Campbell 1969d, 367–77). Speaking from the point of view of humanistic existential socialism, Janousek (1970) has criticized "Reforms as Experiments" (Campbell 1969e) on these grounds. He argues that the whole orientation of assigning persons to treatments by randomization betokens an authoritarian, paternalistic imposition, treating citizens as passive recipients rather than as co-agents directing their own society, treating fellow citizens as "subjects" in the psychologist's and monarchist's sense, as "victims" of the experiment rather than as collegial agents of the experiment.

The enforcement of assigned treatments also violates the egalitarian and voluntaristic ideals of the experimenting society. The disguised experiment violates these too and, in addition, the values of openness, honesty, and accountability. However much we may weight the value

of scientific information in deciding upon the ethics of deceit and lack of informed consent in harmless experimentation done to test scientific theories (Campbell 1969d, 270–77), *social experimentation for policy decisions must adhere meticulously to means idealism on these issues and should include no research procedures that would be excluded as a part of regular governmental procedures.* Participation in policy experiments is more akin to participating in democratic political decision making than to participating in the psychology laboratory. These restrictions all have costs in the validity of experimental inference. They are costs that we must live with and try to compensate for in other ways.

Thus the task of first priority for the methodologists of the experimenting society is to design experimental arrangements that obviate these difficulties. Janousek (1970) has suggested that some system of rotation between the roles of experimenter and subject be designed. While this is not obviously feasible, it is worth more detailed consideration. The following suggestions are likewise only initial fumblings toward possible solutions. For example, experiments using volunteers who are informed of the treatment and control-group alternatives and agree to accept whatever random assignment they draw—such experiments seem to me well worth doing. On some problems, such as public housing versus rental vouchers, they might be as informative as disguised experiments, but even if not, the experimenting society may have to make do with them. The New Jersey Negative Income Tax experiment (Kershaw and Fair 1976; Watts and Rees 1977), the most famous of recent social experiments, participants were told about the particular treatment they were asked to volunteer for, but were not informed of what others would be getting. Should the control group have been told of the thousand-dollar supplemental income it was missing? Perhaps the experimenting society of the future will decide it should have been. Today's methodologists, however, regard the envy and resentment that might thus be generated as too great a threat to experimental validity to be tolerated. Here is a tangle of problems we should be working on right now.

It is my tentative judgment that a variety of random-assignment experiments will be acceptable (see Cook and Campbell 1979, ch. 8). But even more important, we must be elaborating high-quality quasi-experimental approaches that, while more ambiguous in terms of clarity of scientific inference, are usually more acceptable as processes we

will be willing to make a permanent part of our political system. Note, however, that the most common of these quasi-experiments (multivariate regression approaches, covariance in the absence of randomization, path analysis, and causal modeling) have had the bias of making those special programs given to the neediest look mistakenly harmful (Campbell and Boruch 1975e; Cook and Campbell 1979, chs. 3, 4, and 7).

#### *Getting Mutual Criticism and Competitive Replication into Social Experimentation and Program Evaluation*

As my introductory citation of Karl Popper made clear, the objectivity of physical science does not come from the fact that single experiments are done by reputable scientists according to scientific standards. It comes instead from a social process which can be called competitive cross-validation (Campbell 1986a) and from the fact that there are many independent decision makers capable of rerunning an experiment, at least in a theoretically essential form. The resulting dependability of reports (such as it is, and I judge it to usually be high in the physical sciences) comes from a social process rather than from dependence upon the honesty and competence of any single experimenter. Somehow in the social system of science a systematic norm of distrust (Merton's [1973] "organized skepticism") combined with ambitiousness leads people to monitor each other for improved validity. Organized distrust produces trustworthy reports.

This competitive cross-validation will be very hard to achieve, and specific plans must be made to get some semblance of it. To judge from the American experience, government prefers to turn a major pilot study over to a single investigative team. The U.S. Congress is apt to mandate an immediate, nationwide evaluation of a new program to be done by a single evaluator, once and for all, subsequent implementations to go without evaluation. Here are a variety of recommendations to correct this, none of which are practiced in the United States.

*The Contagious Cross-Validation Model for Local Programs.* A generous and concerned government provides funds for developing local programs addressed to a chronic problem. This local program funding includes funds for whatever evaluation the program designers

want, including funds for academic consultants. As a result, there are lots of local programs. When any one of them, after a year or so of debugging, is deemed to be a program others would consider worth borrowing, only at that time would there be program evaluation in a serious sense. Our slogan would be, "Evaluate only proud programs!" (Contrast this with our present ideology, in which Washington planners in Congress and the executive branches design a new program and command immediate nationwide implementation, with no debugging, plus an immediate nationwide evaluation.)

When the high-morale programs and program results were disseminated, there would, no doubt, emerge a group of willing adopters. At this stage, our national funding would support adoptions that included locally designed cross-validating evaluations, including funds for appropriate comparison groups not receiving the treatment. (We might at this or the next stage have large-scale "external" evaluations, as long as these did not preclude interpretable comparisons at each site not depending upon full national implementation.) After five years we might have 100 locally interpretable experiments. We would also have a community of applied social scientists familiar with them all, that had cross-examined each others' data, suggested and done reanalyses, performed bias-sensitive meta-analyses, etc. Many of these scholars would be tenured university or public school faculty whose job security did not depend upon the outcome. From the consensus of this mutually monitoring research community we would advise government and potential adopters.

The contagious cross-validation model is much closer to the model of the physical sciences. Applied social sciences has more, rather than less, need for mutual criticism, argumentative re-analyses, and cross-validation than does physical science because we lack the possibility of experimental isolation, because our data have to be generated through the cooperation of persons with strong stakes in the outcome, and because applied science (either physical or social) is done in an arena in which the rival interests in what the outcome is are so powerful that objective description can become a minor motive.

*Getting Competitive Replication into National Policy Pilot Studies.* The contagious cross-validation model is only appropriate where the program under study can be implemented autonomously by a local unit (be it school, classroom, city, retail store, factory, etc.). Where the

program being piloted has to be eventually implemented nationally, different sources of competitive cross-validation must be sought.

A. Rather than awarding a single contract, each should be *split into two or more independent experiments*, so that all of the hundreds of discretionary decisions as to how to present the experimental treatment and design the questionnaires and interviews would be made and implemented by at least two independent research teams. Such heteromethod replication (Campbell 1969d; Cook and Campbell 1979) is needed for interpretative validity. It would also provide a small group of informed scientists for competitive cross-examination.

B. *Adversarial stake-holder participation* in the design of each pilot experiment or program evaluation and again in the interpretation of results (Krause and Howard 1976; Bryk 1983). We should be consulting with the legislative and administrative opponents of the program as well as the advocates, generating measures of feared undesirable outcomes as well as promised benefits.

C. *Competitive reanalysis* of data from the big studies: the Office of Economic Opportunity created a great precedent we have inadequately lived up to. The Institute for Research on Poverty, University of Wisconsin, has available for reanalysis the data tapes for the New Jersey Negative Income Tax Experiment, and proper scientific disagreements are emerging, for example, as to how they handled the attribution problem (Boeckmann 1981). They have the data from the first big Head Start evaluation, a data-set with a fine record for productive second-guessing (Smith and Bissell 1970; Barnow 1973; Magidson 1977; Bentler and Woodward 1978). Major classics in this area come from my Northwestern University colleagues (Cook et al. 1975; Boruch 1978; Boruch, Wortman, and Cordray, and Associates 1981; Trochim 1982). The original Coleman report (1966) on educational desegregation has been thoroughly reanalyzed, so that now we could assemble a half dozen volumes the size of Mosteller and Moynihan's (1972), and from a modern postpositivist theory of science, we can recognize that only now do we have a competent applied social science community ready to use the Coleman report in conjunction with all related research prior and subsequent to guide governmental policy.

While these secondary analyses are of great value, and should become obligatory for all expensive data collections, we should remember that they cannot fully correct for the hundreds of idiosyncratic discretionary judgments involved in the initial data collection.

D. *Legitimizing dissenting-opinion research reports* from members of the research team. The Freedom of Information Act of the late 1960s was one of the great social inventions increasing the possibility of a valid, policy-relevant, applied social science. While Rights of Subjects legislation (another great innovation) has often been used to greatly curtail its practical implementation (needlessly so—see Campbell, Boruch, Schwartz, and Steinberg 1977d; Boruch and Cecil 1979, 1982; Campbell and Cecil 1982e), the legitimating value is still there. What I propose is that we use such Freedom of Information and right of re-analysis to give to every research assistant on any social research the right to publish independently on the data collected. A background for my argument is the great value that “whistle-blowing” has had for the validity of physical and biological research results when these have been done under conditions of extreme policy relevance. (I am thinking of research on the dangers of chemicals to manufacturing workers and food consumers, the dangers to and effects on humans and sheep of irradiation from nuclear experiments and power generators.) While such whistle-blowing occurs, it is still experienced as a guilt-producing team disloyalty, both by the whistle-blower and coworkers, who may react with ostracism. It would improve the scientific and political validity of applied physics, chemistry, and biology if whistle-blowing were legitimated by reconceptualizing it as the right and duty to generate dissenting-opinion research reports, and if all laboratory staff were provided official access to all data for this purpose. Insofar as our research results are inherently more ambiguous, we are more in need of this in applied social science.

#### *Opinion Surveys as Voting Opportunities*

Opinion surveys would be of central importance in an experimenting society, as “social indicators” of the effects of new programs. Treating opinion surveys as an ideal political decision-making process akin to voting would require great changes. Here are a few: Interviewees would be told who had paid for the questions and how their answers would be used. They would know what programs were being evaluated by their answers (Campbell and Cecil 1982e). They would be given the results of the survey, just as they are given voting results. They would be allowed to use these results in political debates. They would be “co-owners” of the opinions they had created. From the

point of view of present-day social science methodology, these changes would make opinion surveys less valid. Respondents would distort their opinions in deliberate efforts to influence governmental decisions as to which programs should be continued or which regions were most in need of more resources. There would even be campaigns to get people to answer interviews in certain ways.

These costs in “validity” are probably unavoidable. Opinion surveys, however, would still be useful and informative, just as are votes now, once we got used to these new conditions of meaning. The methodological problems involved are ones we should be working on and indeed, are ones best researched in the transition period.

#### *The Corrupting Effect of Using Social Science Indicators*

The social indicators that are now being used include public records as well as opinion surveys—records of deaths, diseases, crimes, accidents, incomes, and school-achievement test scores. In the experimenting society of the future, these would probably be used even more as indicators of how programs are doing. It might be thought such records would be more resistant to bias than are interview data. More resistant, yes, but still subject to a discouraging law that seems to be emerging: the more any social indicator is used for social decision making, the greater the corruption pressures upon it (Campbell 1979a, 84–86; Ginsberg 1984). Measures that have been valid for describing the state of society become invalid when they start being used for political decision making. Moreover, such use often leads to a destructive corruption of the social process that the indicator was designed to measure. Thus, the U.S. war on crime of 1968–71 achieved its results by downgrading the seriousness of crimes as recorded. Thus scoring police departments for their efficiency in solving crimes, combined with plea bargaining, has led to a situation where the burglar who confesses to the most unsolved crimes is given the lightest sentence. Thus achievement tests once valid for describing educational status have become less valid when used as the basis of rewards to students or teachers. Thus in the Russian five-year plans, setting quantitative quotas for factories produced wasteful distortions of production, as in the nail factories that overproduced large spikes when the quota was set by tonnage, and overproduced small nails when the quota was set by number of items turned out. Thus, the U.S. Army

reform in reporting enemy casualties in terms of body counts rather than subjective estimates, and in using body counts to evaluate the effectiveness of field commands, created an immoral and irrelevant military goal in Vietnam. Thus census data for Chicago are more trustworthy than voting records, not because the census-taking is more protected against fraud—quite the contrary—but because the voting data are used so much more in political decision making.

In the experimenting society, such social indicators will be used more than they are at present, and the corruption pressures will thus be greater. This problem seems to me so serious that, while I believe it is solvable, it provides one more reason why we should not rush forward into the experimenting society until it is solved. We need social-system inventions and studies of how some high-pressure indicators, such as the cost-of-living index, remain so free of distortion. My tentative solution to this problem is to distinguish carefully between two movements in "scientific government." I end up opposing the use of quantitative indicators for achieving managerial control, the accountability movement. The regularized use of such measures and the focus on evaluating specific social units and their administrators seems to me to create more evils than it cures. The other movement is program evaluation. As I have expressed under "Resistance to Assessing Outcomes" above, the temporary use of quantitative measures in evaluating alternative programs, which existing staffs could implement, seems to me still beneficial to the experimenting society.

#### *Integrating the Evidence into Political Decision*

One solution to bias is to use multiple indicators of the same problem, each of the indicators being recognized as imperfect, but so chosen as to have different imperfections, different susceptibilities to distortion. This produces a variety of estimates of program effectiveness, benefits, and problems. The judgmental task of pooling all of these indicators is one more appropriate for an elected legislature than for a committee of social scientists. Furthermore, in an experimenting society we would be doing scientific evaluations on many more programs than we are doing now. With multiple measures on multiple programs we will have created a monster of measurement, a formidable information overload. How to reconcile our need for facts with democratic decision making is another problem we must solve before we wel-

come the experimenting society. New institutions will be needed, such as an auxiliary legislature of quantitative social scientists, each appointed by one real legislator. Such an auxiliary legislature could process advisory decisions, with full attention to the scientific evidence. The real legislator could then guide his own decisive vote by the auxiliary legislature's actions and the issues raised. This awkward and expensive procedure seems to me better than delegating this process to an appointed scientific elite.

I offer this unwieldy suggestion to testify to my estimate of the seriousness of the problem and to emphasize our need for creative speculation as to how it might be solved. Such conjectures can right now be exposed to vigorous criticism. The ones we judge most plausible we will want to try out in practice. After implementing them—or even before—we may decide that they remain so unsolved as to be a reason to withhold our advocacy of the experimenting society.

#### *Legitimizing and Facilitating Evaluation by Nonprofessional Participants and Observers*

We applied social scientists, methodological servants of the experimenting society, are like any profession (see Ivan Illich) in danger of becoming a self-serving elite, in whose professional interest it is to make program evaluation an esoteric art requiring complex statistical procedures that make our conclusions immune from criticism, even from well-placed, competent observers who saw the program in action. We are also apt to become unwittingly co-opted into a pervasive bias in favor of the already-established governmental and extra-governmental powers, who, after all, will usually be the source of our past and future salaries.

To avoid such biases, we must devise ways that are readily comprehensible to the participating staffs, recipients, and other well-placed observers for them to collect, formulate, and summarize their estimations of program effectiveness (Campbell 1978b). We must recognize that such summaries may have a validity comparable to the statistical analysis of more formal measures. Usually these perspectives will agree, but where they do not, we should remember that the statistical analyses involve simplifying assumptions that may be seriously in error (Campbell and Erlebacher 1970c; Campbell and Boruch 1975e).

We must also remember that in social experimentation, the lack of

controlled conditions makes necessary the technique of explicitly developing and evaluating "threats to validity" and "plausible rival hypotheses" (Campbell and Stanley 1963c; Cook and Campbell 1979). For this purpose, it is those who have situation-specific information who make the best critics, and the best judges, of the plausibility of most of the rival hypotheses in their specific setting. We must develop procedures for eliciting such criticisms and judgments.

In this process, we must provide these nonprofessional observers with the self-confidence and opportunity to publicly disagree with the conclusions of the professional applied social scientists.

#### *Long-term Follow-up*

The relevant efficacy of social experiments is long term, the effect in subsequent decades (rather than, or in addition to, months or years). For experiments involving children (compensatory education, preventive interventions, kibbutz child rearing), the effects on their adult lives, fifteen to twenty years after the experiment, are most relevant. Such long-delayed evaluations are jeopardized by several different classes of problems.

1. Locating and assessing experimental participants twenty years later presents formidable problems. (Some U.S. compensatory-education efforts have had 50 percent failures in efforts to recontact and remeasure, after only one year.)
2. The scientists doing research on the initial intervention are unlikely to still have that research focus ten to twenty years later. Their focal scientific community for the experimental problem will also have dispersed.
3. The governmental initiative funding the research will have dissipated; new focal concerns will have higher priority.

Thus it is essential to have immediate outcome measures, focused on proximal indicators of those traits that the theory involved specifies as mediating the long-term benefits. But since the theories involved are not themselves well tested, and indeed, can only be tested by long-term follow-ups, it is a major responsibility to prepare the basis for later follow-ups.

*Snowballing versus dissipating effects.* One should distinguish in

anticipation between two contrasting forms of impact for interventions delivered at a specific time: snowballing versus dissipating, and the existence for most interventions of an unknown threshold point between them. Consider two randomized experiments using samples of 100 college seniors (with blocking on parental wealth, ability, and undergraduate major). For one experiment, members of the treatment group are each given \$5,000, their controls nothing. It might be anticipated that in each subsequent year the difference between the net worth of the experimentals and controls would diminish until five years later no significant difference would remain. On the other hand, were the experimental treatment to be \$50,000, a snowballing effect might be found, in which the differences between the experimental and control group steadily increased in subsequent years.

For many interventions, for example those that leave children in the same environment and involve small percentages of their waking hours, we must expect dissipating effects, unless a particularly sensitive "imprinting" time has been hit. The laws of learning, interference, forgetting, and spontaneous recovery provide theoretical support for this prediction (Campbell and Frey 1970d). Think of the multiplicity of change agents, other than the intervention, that impinge upon the outcome variable in one direction or another. These collectively produce a random-like scatter of effects, more numerous the more time has elapsed. Relative to the sum total of these effects, the intervention is but a drop in the bucket. Moreover, as time goes on, weights of pre-intervention influences increase relative to that of the intervention when it first occurred. (Such a theory argues for small dosage, short-term experiments to winnow out a few promising treatments. In later experiments these would be applied intensively over many years, looking for a long-term effect.)

But such theory can lead to undue pessimism. Several recent reports on long-term follow-ups of preschool compensatory education show lasting effects (Lazar and Darlington 1982; The Consortium for Longitudinal Studies 1983; Weikart et al. 1984; Crain et al. 1984). For cognitive effects as measured by achievement tests, the dissipating pattern has been found (considering these as akin to vocabulary tests makes the learning-theory expectations more obvious). But for academic self-confidence, persistence in seeking higher education, and avoidance of delinquency, impressive impacts have been found fifteen years later. Note that on these variables no immediate proximal media-

tor was effectively measured. Making decisions on long-term follow-ups based upon test scores received five years later would have precluded discovery of the important long-term effects. Many exploratory studies will validly be deemed unworthy of long-term follow-ups, based upon the experimental pilot studies employing short-term outcome measures. But if the costs of archiving the records that would make long-term follow-ups possible is not too great, I recommend routine recording of such data, because later theoretical reconsiderations may lead to changes in such decisions.

Another classic in long-term follow-up on a deliberate experimental intervention is McCord's (1978, 1981) thirty-year follow-up on the Cambridge-Somerville delinquency prevention project. Here again, very significant long-term impact was documented, but in this case, unanticipated harmful effects the opposite of those intended. Caplan (1968), for the Chicago Boys Club study, has likewise found what appear to be harmful effects from an extensive gang-worker program, in a five-year follow-up study. The methodological difficulties of long-term follow-ups are such that one might well be skeptical of the outcomes claimed, whether beneficial or harmful. So far as we are aware, a thorough scrutiny of such problems and the plausible alternative explanations of the outcomes they provide has not yet been undertaken. Differential attrition for experimentals versus controls during treatment and differential rates of locating cases at follow-up time are just two of the possibilities. We have major responsibility to make a profound cross-examination of these studies and others, both for the intrinsic interest in their outcomes and, more generally, for their methodological implications in preparing plans for possible long-term follow-ups.

*A survey of methodological problems and solutions in long-term follow-up studies.* We propose a methodological review in this area. As a beginning, one might make extended visits to scholars and centers having such experience, such as those cited above, but also to include those whose follow-ups are not centered on deliberate experimental interventions, but are rather correlational or are impact studies of exogenous "treatments," such as the Great Depression of the 1930s, parental employment, parental death, mode of child rearing, etc. A major center for such studies is in the Institute for Human Development at the University of California, where follow-ups are still being conducted on longitudinal study cases started in the early 1930s by the

Institute of Child Welfare under grants from the Laura Spellman Rockefeller Foundation. At least three studies are involved; two beginning at birth or before ("The Guidance Study," initiated by Jean Walker Macfarlane, and "The Growth Study," under Nancy Bailey) and one, "The Adolescent Study," by Harold E. Jones, Mary Cover Jones et al.) beginning with fifth-graders. The Fels Institute of Yellow Springs, Ohio, also started longitudinal studies of infants and families in the 1930s with some follow-ups into the late 1940s at least.

This methodological review should include not only the experimental threats to validity, including the massive "pretest" effects, but also should pay attention to sociology-of-science issues in regard to what kinds of institutional arrangements and/or accidents of careers made it possible to carry out repeated follow-ups (as in the University of California Berkeley group) or follow-ups after long periods of no contact with the participants (as in the McCord [1978] classic, the Crain [1984] study, and some of those in the cooperating group of Lazar and Darlington [1982]). Even such issues as who provided the documentation, storage space, and retrieval competence in the archiving are important. Analysis of whether current legislation and custom on privacy and other rights of subjects would have permitted such follow-up would also be germane.

Other stages of this survey might be achieved by contracted papers and conferences, bringing together those who have done such follow-ups or have planned them. In the latter category, note that in the Denver-Seattle Negative Income Tax Experiment, being coordinated by the Institute for Research on Poverty of the University of Wisconsin, Madison, a small subsample has been guaranteed a minimum income for a twenty-year period to allay economists' worry that the announced three-year intervention in the New Jersey studies (Kershaw and Fair 1976; Watts and Rees 1977) artificially produced a high sustained work effort. Presumably, appropriate controls will be followed up also.

*Guidelines for preparing for long-term follow-up.* Subject to revision after the study described above, the following temporary recommendations are offered:

1. Names, addresses, parents' names, place of birth, and social security numbers should be recorded and archived under confidentiality-preserving conditions for all designated experimental and control partici-

- pants.
2. Informed consent should include mention of the possibility of long-term follow-up. Institutional Review Board approval need not be obtained for the follow-up since that will not be part of the research currently funded. However, the importance of possible later long-term follow-ups should be used as justification to the I.R.B. for the retention of identifying information facilitating such follow-up.
  3. Identifying information and reason for loss of each case should be retained for all types of sample attrition from the very first designation of target populations. (Since many of the potential long-term outcomes will have no appropriate "pretest," outcomes for all those randomly assigned, even if lost before any pretesting or treatment, will have to be analyzed to estimate the degree to which selection bias through differential attrition explains the results [Cook and Campbell 1979, ch. 8].)
  4. Feasibility of long-term follow-up should be one criterion in selecting among research projects to be supported. (Particularly attractive are "encompassing measurement frameworks" that will provide later indicators for those who fall away due to attrition as well as those who remain within the experimental and control samples. School records, earnings subject to withholding tax, intake records from mental hospitals and the criminal justice bureaucracy provide such.) Sometimes a sample will be chosen just because the experimental and control samples can be from a relevant measurement framework. Thus a study might be centered upon life insurance salesmen just because sales volume was a relevant measure of effective adjustment. (A little thought should provide better examples than this.)

*Strategies for achieving continuity necessary for long-term follow-up.* We have already noted that the need new members of Congress and new administrators have to introduce innovations that can be credited to them works against project longevity. We can also note that the need for news of prompt "breakthroughs" on the part of those legislators and administrators who took political risks to establish and maintain a given experiment makes even these founding supporters reluctant to put emphasis on pilot studies that won't pay off until fifteen years later. An adequate, hypothetically normative sociology of scientific validity would address this problem and provide theoretical understandings that could suggest remedies. We are not at that stage yet, but we should begin brainstorming on the problem.

Educating our sponsors to the problem, to their predicaments as well as to ours, and to the destructive effects of some understandable reactions to their predicaments as well as to ours, and to the destructive effects of some understandable reactions to their predicaments

might be tried. Once pointed out, and with the compensatory education results as illustrations, the necessity for long-term follow-up should be obvious.

In our dealings with our patrons (and in their dealings with their constituencies), it may help to emphasize the importance of the problem and how that importance justifies our trying out many approaches to see what works, rather than falling into the tempting "overadvocacy trap" (Campbell 1969e) in which we (and our patrons) promise too much for specific new "solutions" just in order to get them tried out at all. This strategy has been criticized (Shaver and Staines 1971; Campbell 1971d) at the level of treatment program leadership, but the disadvantages alleged would not seem to hold for many experimental interventions.

#### *The Experimenting Society as Normal Rather than Extraordinary or Revolutionary Science*

As a final issue, let me confess or claim that in terms of Thomas Kuhn's (1970) still influential dichotomy, the experimenting society represents normal rather than revolutionary science. Popper can be quoted two ways on this. As a depiction of science (Popper 1970), he has deprecated normal science and called for the ideal of continual revolution. But in his model for the experimenting society, he has asserted: "A social technology is needed whose results can be tested by piecemeal social engineering." This is what I am classifying as a normal science model for an experimenting society, in Kuhn's category system.

According to Kuhn, there are normal periods of scientific growth during which there is general consensus on the rules for deciding which theory is more valid. In contrast, there are extraordinary or revolutionary periods in science, in which the choices facing scientists have to be made on the basis of decision rules which are not a part of the old paradigm. Initially, the choice of the new dominant theory after such a scientific revolution is unjustified in terms of the decision rules of the prior period of normal science.

For social experimentation, the Kuhnian metaphor of revolution can be returned to the political scene. Evaluation research is clearly something done by, or at least tolerated by, a government in power. It presumes a stable social system generating social indicators that re-

main relatively constant in meaning so that they can be used to measure the program's impact. The programs that are implemented must be small enough not to seriously disturb the encompassing social measurement system. Thus the technology I have been discussing is not available to measure the social impact of a revolution. Even within a stable political continuity, it may be limited to relatively minor innovations. It presupposes a stable society with governmental stability both in general and particularly in record keeping. It presupposes the meaningful comparison of present, past, and future. The experimenting proposed here is within the framework of such a society, any given experiment being of such a small magnitude as to not fundamentally change that society. For any given step, it is limited to small changes rather than fundamental framework changes.

We gradualists will argue that fundamental changes can be made in this way, that by small steps, each validated as improvements, we can move toward any optimal society. The revolutionist will agree with Kuhn that the framework is the problem, not the details, that radical changes going beyond the framework are needed and that these are not likely to be judged scientifically by the criteria of the preexisting framework.

For intentional social changes at any level, I would argue that we should not hold up implementing them, just because their impacts cannot be measured. A trivial example that comes to mind is the shift in the U.S. public schools some twenty years ago from traditional teaching of arithmetic to the "new math" based upon set theory. Since the "new math" made the available achievement tests and records based upon old math inappropriate, the superiority of "new math" had to be taken on the basis of consensus of expert judgment, rather than pilot studies demonstrating its superiority. (Now, twenty years later, we could do rather elegant quasi-experimental evaluations of the impact of new math on mathematics courses taken at the university level and the quality of performance in them. This possibility exists because the innovation was decided upon by local school districts, at a variety of dates over several years, rather than being implemented simultaneously.)

Extending this reasoning, I will concede that it is not a crucial argument against political revolution to say that it may destroy the measurement series by which the revolution's impact might be precisely measured. If there be a net argument against revolutions over

cumulative, gradual, tested-out, step-wise change, it has to be made on the basis of other side effects. After Popper's *The Open Society and its Enemies* ([1945], 1952) and his *Poverty of Historicism* (1944), we can no longer credit revolutionary programs with superior scientific status over reformist ones. But while revolution may not provide in itself an evaluable social experiment, as a new status quo, just as an old status quo, it could be followed by an experimental approach to fine tuning its praxis in order to achieve the optimal implementation of the ideals and idealism that brought it into being.

### Concluding Comments

We have gone over a number of the problems that have to be solved before we can wholeheartedly advocate an experimenting society. While it is clear on many counts that a totalitarian society is not likely to be an experimenting society, it is also clear that it holds many problems for an open society.

For now, let a number of us become methodologists for the experimenting society, remembering that as we develop in detail the procedures, possibilities, and problems of the experimenting society, we will be acquainting ourselves with what it would be like as well as this can be done in advance. As this portrait emerges in greater clarity, it will be our duty to continually ask ourselves if we really want to advocate this monster of measurement and experimentation. We must share the developing picture with the most articulate and hostile critics of such a society and consider in detail their warnings.

If it is not a future we want, who should know better or sooner than we, the ambivalent methodologists of the experimenting society.

### References

- Barnow, B.S. 1973. The Effects of Head Start and Socioeconomic Status of Cognitive Development of Disadvantaged Children. Ph.D. Dissertation. University of Wisconsin, Department of Economics.
- Boeckman, M.E. 1981. "Rethinking the Results of a Negative Income Tax Experiment." In *Reanalyzing Program Evaluations*, edited by R.F. Boruch, P.M. Wortman, and D.S. Cordray. 341–63. San Francisco: Jossey-Bass.
- Boruch, R.F. 1978. *Secondary Analysis*. San Francisco: Jossey-Bass.
- Boruch, R.F. and J.S. Cecil. 1979. *Assuring the Confidentiality of Social Research Data*. Philadelphia: University of Pennsylvania Press.
- Boruch, R.F., Wortman, D.F., Cordray, D.S. and associates. 1981. *Reanalyzing Program Evaluations*. San Francisco: Jossey-Bass.

- Caplan, Nathan. 1968. "Treatment Intervention and Reciprocal Interaction Effects." *Journal of Social Issues* 24, 1: 63–88.
- Campbell, Donald T. 1957. "Factors Relevant to the Validity of Experiments in Social Settings." *Psychological Bulletin* 54: 297–312.
- Campbell, Donald T. 1969a. "Definitional versus Multiple Operationalism." 2: 14–17, *SP*, 31–36.
- Campbell, Donald T. 1969b. "Ethnocentrism of Disciplines and the Fish-scale Model of Omniscience." In M. Sherif and C.W. Sherif, eds., *Interdisciplinary Relationships in the Social Sciences*. 328–48. Hawthorne, NY: Aldine.
- Campbell, Donald T. 1969c. "A Phenomenology of the Other One: Corrigible, Hypothetical, and Critical." In T. Mischel, ed., *Human Action: Conceptual and Empirical Issues*. 41–69. New York: Academic Press.
- Campbell, Donald T. 1969d. "Prospective: Artifact and Control." In R. Rosenthal and R. Rosnow, eds., *Artifact in Behavior Research*. 351–82. New York: Academic Press.
- Campbell, Donald T. 1969e. "Reforms as Experiments." *American Psychologist* 24: 409–29.
- Campbell, Donald T. 1971. "Comments on the Comments by Shaver and Staines." *Urban Affairs Quarterly* 7,2: 187–92.
- Campbell, Donald T. 1974. "Evolutionary Epistemology." In *The Philosophy of Karl Popper*, edited by P.A. Schilpp, 413–63. LaSalle, IL: Open Court Press.
- Campbell, Donald T. 1978. "Qualitative Knowing in Action Research." In *The Social Context of Method*, edited by M. Brenner, P. Marsh, and M. Brenner, 184–209. London: Croom Helm.
- Campbell, Donald T. 1979. "Assessing the Impact of Planned Social Change." *Evaluation and Program Planning* 2: 67–90.
- Campbell, Donald T. 1986b. "Science's Social System of Validity-Enhancing Collective Belief Change and the Problems of the Social Sciences." In *Metatheory and Social Science: Pluralisms and Subjectivities*, edited by D.W. Fiske and R.A. Schweder, 108–35.
- Campbell, Donald T. and Robert F. Boruch. 1975. "Making the Case for Randomized Assignment to Treatments by Considering the Alternatives: Six Ways in Which Quasi-Experimental Evaluations in Compensatory Education Tend to Underestimate Effects." In *Evaluation and Experiment: Some Critical Issues in Assessing Social Programs*, edited by C.A. Bennett and A. Lumsdaine, 195–296. New York: Academic Press.
- Campbell, Donald T., Boruch, R.F., Schwarz, R.D. and J. Steinberg. 1977. "Confidentiality—Preserving Models of Access to Files and to Interfile Exchange for Useful Statistical Analysis." *Evaluation Quarterly* 1: 269–99.
- Campbell, Donald T. and J.S. Cecil. 1982. "A Proposed System of Regulation for the Protection of Participants in Low-Risk Areas of Applied Social Research." In *The Ethics of Social Research: Fieldwork, Regulation, and Publication*, edited by J.E. Sieber, 96–121. New York: Springer-Verlag.
- Campbell, Donald T. and A. Erlebacher. 1970. "How Regression Artifacts in Quasi-Experimental Evaluations Can Mistakenly Make Compensatory Education Look Harmful." In *Compensatory Education: A National Debate*, edited by J. Hellmuth, vol. 3: *Disadvantaged Child*, 185–225. New York: Brunner/Mazel.
- Campbell, Donald T. and P.W. Frey. 1970. "The Implications of Learning Theory for the Fade-Out of Gains From Compensatory Education." In *Compensatory Education: A National Debate*, edited by J. Hellmuth, vol. 3: *Disadvantaged Child*, 445–63. New York: Brunner/Mazel.

- Campbell, Donald T. and Julian C. Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.
- Coleman, J.S. et al. 1960. *Equality of Educational Opportunity*. Washington, D.C.: Office of Education, U.S. Department of Health, Education, and Welfare.
- Consortium for Longitudinal Studies. 1983. *As the Twig is Bent....Lasting Effects of Preschool Programs*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cook, T. D., et al. 1975. *Sesame Street Revisited: A Case Study in Evaluation Research*. New York: Russell Sage Foundation.
- Cook, Thomas D. and Donald T. Campbell. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston, MA: Houghton Mifflin.
- Dunn, E.S. 1971. *Economic and Social Development: A Process of Social Learning*. Baltimore: The Johns Hopkins University Press.
- Etzioni, A. 1968. *The Active Society*. New York: The Free Press.
- Ginsberg, P.E. 1984. "The Dysfunctional Side-Effects of Quantitative Indicator Production: Illustrations from Mental Health Care." *Evaluation and Program Planning* 7: 1–12.
- Haworth, L. 1960. "The Experimental Society: Dewey and Jordan." *Ethics* 71, 1: 27–40.
- Janousek, J. 1970. "Comments on Campbell's 'Reforms as Experiments.'" *American Psychologist* 25: 191–93.
- Kershaw, D. and J. Falr. 1976. *The New Jersey Income Maintenance Experiment. Volume I: Operations, Surveys, and Administration*. New York: Academic Press.
- Kuhn, T.S. 1970. *The Structure of Scientific Revolutions*, 2nd ed., enlarged. Chicago: University of Chicago Press.
- Lazar, I. and R. Darlington. 1982. "Lasting Effects of Early Education: A Report from the Consortium for Longitudinal Studies." Monographs of the Society for Research in Child Development 47: 2–3.
- Lewin, K. 1946. "Action Research and Minority Problems." *Journal of Social Issues* 2: 34–46.
- Lewin, K. 1948. *Resolving Social Conflict*. New York: Harper.
- Magidson, J. 1977. "Toward a Causal Model Approach for Adjusting Pre-Existing Differences in the Nonequivalent Control Group Situation." *Evaluation Quarterly* 1, 3: 399–420.
- McCord, J. 1978. "A Thirty-Year Follow-Up of Treatment Effects." *American Psychologist* 33: 284–89.
- McCord, J. 1981. "Considerations of Some Effects of a Counseling Program." In *New Directions in the Rehabilitation of Criminal Offenders*, edited by S.E. Martin, L.B. Sechrest, and R. Redner, 393–405. Washington, DC: National Academy Press.
- Merton, Robert K. 1973. *The Sociology of Science*, edited by N.W. Storer. Chicago, IL: The University of Chicago Press.
- Mosteller, F. and D.P. Moynihan (eds.) 1972. *On Equality of Educational Opportunity*. New York: Vintage Books.
- Polanyi, M. 1966. "The Message of the Hungarian Revolution." *American Scholar* 35: 261–76.
- Popper, K.R. 1944. *The Poverty of Historicism*. London: Routledge & Kegan Paul.
- Popper, K.R. 1952. *The Open Society and Its Enemies*. London: Routledge & Kegan Paul.
- Popper, K.R. 1970. "Normal Science and Its Dangers." In *Criticism and the Growth of Knowledge*, edited by I. Lakatos and A. Musgrave, 51–58. Cambridge: Cambridge University Press.

- Ravetz, J. 1971. *Scientific Knowledge and Its Social Problems*. Oxford: Clarendon Press.
- Rossi, P.H. 1969. "Practice, Method, and Theory in Evaluating Social-Action Programs." In *On Fighting Poverty*, edited by J. L. Sundquist, 217–35. New York: Basic Books.
- Sanford, N. 1970. "What Ever Happened to Action Research?" *Journal of Social Issues* 26: 2–23.
- Schwarzlander, H. 1978. "The Process Utopia." *Alternative Futures* 1,2: 99–102.
- Shaver, P. and G. Staines. 1971. "Problems Facing Campbell's 'Experimenting Society.'" *Urban Affairs Quarterly* 7,2: 173–86.
- Smith, M.S. and J.S. Bissell. 1970. "Report Analysis: The Impact of Head Start." *Harvard Educational Review* 40: 51–104.
- Trochim, W. 1982. "Methodologically Based Discrepancies in Compensatory Education Evaluations." *Evaluation Review* 6: 4–5.
- Watts, H. W. and A. Rees (eds.) 1977. *The New Jersey Income Maintenance Experiments*. Volume 2: *Labor-Supply Responses*. Volume 3: *Expenditures, Health and Social Behavior, and the Quality of the Evidence*. New York: Academic Press.
- Weikart, D.P., Berrueta-Clement, J.R., Schweinhart, L.J., Barnett, W.S., and A.S. Epstein. 1984. *Changed Lives: The Effects of the Perry Preschool Program on Youths Through Age 19*. Ypsilanti, MI: High/Scope Press.

## 3

## Making Science into an Experimenting Society

*Steve Fuller*

Science certainly acquires knowledge through experimentation, but it is probably the last institution to which one would turn to find the "experimenting society" at work. Perhaps this is why science policy is the least developed area of public policy. In any case, the paradox helps explain public resistance to Campbell's (1988) vision. To regard local communities as so many sites for testing the relative efficacy of treatments or the relative plausibility of hypotheses would seem to require no less than the technocratic colonization of the life-world. Yes, the public is swept up into the scientific enterprise—but as citizen-subjects, not citizen-experimenters. Even in Campbell's canonical presentation of the experimenting society, the public does not initiate research but merely reacts or adapts to it—albeit perhaps in quite critical and sophisticated ways.

Symptomatic of this asymmetry between *experimental science* and the *experimenting society* is the following admonition of Campbell's (1988, 301): "Accepting the important role of social experimentation at once puts clarity of scientific inference into conflict with means-idealism and acceptable political processes." While perfectly reasonable on the surface, this sentence makes two assumptions that invite even more asymmetrical relationships to develop between scientists and the public.

First, it assumes that "the clarity of scientific inference" is corruptible by the introduction of nonscientists or nonscientific considerations.