On Behalf of an Experimental Political Science

Donald R. Kinder and Thomas R. Palfrey

Experimentation should be part of the political scientist's everyday empirical repertoire. For a variety of reasons, none of which has much to do with science, it is not, and political science is poorer for it. In political science, experiments are often regarded as exotic or irrelevant, and experimentation is a subject not quite fit for serious discussion. Experiments are what chemists do or, closer to home, what psychologists do, but not what we political scientists do. The science of politics, so runs the standard argument, cannot be an experimental one.

We think this is wrong. Granted, experiments will never dominate the study of politics, nor should they. But, while an exclusively experimental political science is neither realistic nor desirable, a political science based on a variety of empirical methods, *experimentation prominent among them,* is both within our reach and well worth reaching for, a point we hope to establish here.

With that goal in mind, our first order of business is to argue that the empirical study of politics today rests on too narrow a methodological foundation. Indeed, to the typical student two years into a graduate program in political science, "methods" often means a course or two in statistics and perhaps one devoted to survey research; data collection means placing an order for a tape from the Inter-University Consortium for Political and Social Research. For this regrettable and unhealthy condition, we prescribe a dose of experimentation. Of course, experimentation is no magic elixir: turning to experiments will hardly cure political science of all its troubles. Our argument is that experiments must supplement, not replace, traditional empirical methods. Experimentation is by no means a complete remedy, but it can add a badly needed and valuable dimension to the study of politics—or so we claim. Making such a claim requires us first to define and defend what we mean by experiment, which is the chief business of section 2. There we offer an ecumenical conception, one that reflects both Kinder's training and interest in political psychology and Palfrey's training and interest in political econny. With this definition in place, we go on, in section 3, to develop experi-entation's considerable comparative advantages, and then, in section 4 (and the interest of fairness), to acknowledge experimental disadvantages. In the ssay's final section, we deliver a mercifully brief exhortation on behalf of a ore vigorously experimental political science, and then introduce and orga-ize the readings that follow.

Our aim throughout is the unabashed promotion of experimentation. We ope that experiments will become a more prominent feature of the study of olitics: that more and better experiments will be undertaken, that more and etter students will be trained to do experiments, and, more generally, that the evel of discourse in political science about experimental matters will be elevated.[1] We also hope to nudge those already enlisted under the experimen-tal banner to take notice of experimental research going on within political science outside their special sphere of interest. In this last aim, we hope to provoke the kind of learning in our readers that we have enjoyed in collaborat-ing on this essay and book: a collaboration between two wayward political scientists with very different backgrounds and substantive interests but agree-ing about the power and promise of experimentation.

1. Domination and Diversification in Empirical Methods

In 1924, the question occurred to Harold Gosnell, as it had no doubt occurred to many observers of politics, whether turnout on election day could be enhanced by providing prospective voters with information about registration procedures and by encouraging them to vote. The question was common, but what Gosnell did about it, particularly for his time, was not: he undertook an experiment. In the summer preceding the otherwise forgettable presidential contest between Calvin Coolidge and John W. Davis, Gosnell assigned neigh-borhoods lying within 12 typical districts in the city of Chicago to one of two conditions (or "treatments"). Residents living in neighborhoods designated as experimental were sent postcards that pointed out voter registration deadlines and locations and went on to suggest that citizens of Chicago who failed to exercise their sacred right to vote were little different from those "slackers" who refused to defend their country in time of war. Meanwhile, residents of comparable neighborhoods, assigned to the control condition, were left alone. On election day, about 8 percent more of the experimental group than the control group actually turned out to vote. The answer to Gosnell's question was yes (Gosnell 1927).

1. In this way, we are continuing a line of argument initiated by Brody and Brownstein (1975).

In the more than sixty years that have passed, relatively few political scientists have followed Gosnell's excellent example. Most of what political science does in the name of science has nothing to do with experimentation. Other methods dominate the contemporary scene. Perhaps most of all these days, political scientists with a special interest in American politics do some version of survey research, as in Converse's (1964) seminal analysis of belief systems of the American mass public, based on standardized interviews with large samples of Americans of voting age, or Lane's (1962) intensive conver-sations with a handful of New Haven working-class men, or Walker's (1983) survey of political organizations.[2] Or they do archival research, taking advan-tage of recorded or reconstructed data that index naturally occurring political, social, and economic events. Exemplary illustrations of this approach include Key's examination of aggregate voting returns in *Southern Politics* (1949), Kramer's (1971) influential analysis of the impact of economic fluctuations on U.S. congressional elections in this century, and the rich and controversial literature that has grown up around the timing and meaning of the realignment of the American party system (e.g., Burnham 1970; Clubb, Flanigan, and Zingale 1980; Key 1955; Sundquist 1973). Less often, political scientists undertake simulations of political processes, as in Axelrod's (1984) examina-tion of the evolution of cooperation, or they engage in systematic observation, the virtues of which are illustrated most conspicuously by Fenno's *Home Style* (1978), an eye-opening account of the behavior of House members in their districts. Occasionally—not as often as we would like—they do experiments.

Survey and archival approaches dominate the contemporary political scientist's methodological world. An unsettlingly high proportion of what we think we know—of what we count as dependable knowledge about politics—comes from just these two ways of securing evidence. This state of affairs is unsettling not because survey and archival methods are especially prone to bias and error. Our quarrel is not with survey and archival methods as such, but with their preeminence. A science of politics erected upon such a narrow methodological foundation stands precariously.

The preoccupation with survey and archival methods brings troubles of two sorts in particular. The first is strictly methodological. We assume that *all* methods, including the experimental method that we so admire, are fallible; none provides a royal road to truth. The proper response to this inescapable predicament is to pursue research questions from a variety of methodological angles, all of them fallible, but fallible in different ways. Dependable knowl-edge has its base in no single method, but rather in triangulation across multiple methods. According to Webb et al. (1966)

2. More precisely, political scientists most often undertake *secondary* analysis of survey data: they analyze survey data collected and distributed by others.

> When a hypothesis can survive the confrontation of a series of complementary methods of testing, it contains a degree of validity unattainable by one tested within the more constricted framework of a single method. Findings from this latter approach must always be subject to the suspicion that they are method-bound: Will the comparison totter when exposed to an equally prudent but different testing method? (Webb et al., 1966, 174)

Thus the descriptions and theories that emerge from a political science dominated by survey and archival methods should be greeted with the specific suspicion that they are tied up with a particular and, in some degree, parochial way of examining the political world. Descriptions and theories that survive multiple tests that are based on multiple methods sweep that suspicion aside. As a matter of principle, political science would be wise to diversify its methodological portfolio—and, as we will argue later, there are excellent reasons to diversify in the direction of experimentation in particular.

A second trouble associated with political science's preoccupation with survey and archival methods is more insidious and has to do with the range of empirical questions that political science pursues. Methodological preoccupations are inevitably accompanied by theoretical ones. What is worth doing is seen through the filter of what one is able to do. Only certain questions are deemed interesting or relevant: in the case of political science, they are the questions that allow survey and archival methods room to operate. In the meantime, other questions languish, set aside or never noticed in the first place. This is the "law of the instrument" (Kaplan 1964, 28) at work.

> Give a small boy a hammer, and he will find that everything he encounters needs pounding. It comes as no particular surprise to discover that a scientist formulates problems in a way which requires for their solution just those techniques in which he himself is especially skilled.[3]

The danger here is that one method will come to be thought of as *the* method. Methods that prevail for a time may come to be identified with the scientific method as such and, in the process, become automatic habits. Meanwhile, other methods get crowded out or, as Kaplan (1964, 29) put it, "denied the name of science." There are, of course, real benefits to specializa-

3. We let the male pronouns in the quoted passage stand, not because we believe that female political scientists are immune to the law of the instrument, but because, as we began to write this essay, our sons Benjamin and Ross were deep in the business of ravaging our households with hammers, and, one memorable evening, each other. To us, the picture of the small *boy* hammering is compelling.

tion, but automatic reliance on one method may undermine the imaginative spirit that is essential to scientific advance.

Just as serious, relying on a single method leads inevitably to the definition of some problems as proper and legitimate, while other lines of inquiry are closed off. A good case in point is provided by the rise to prominence of the survey method in studies of elections. Once the exclusive province of a handful of souls undaunted by aggregate voting statistics, the study of elections became dominated by survey researchers within a generation. As exemplified by *The American Voter,* published in 1960, election research came to focus on understanding the social psychological underpinnings of choices made by individual voters. This is a valuable way of looking at elections, but, as survey research became preeminent, other important aspects of elections receded into the background. Put in a stylized way, attention was directed primarily at understanding *voters* as individual decision makers rather than *elections* as collective choices. With a few important exceptions, research found it difficult to pursue other sorts of questions of high relevance to political science. Empirical work that attempted to use election survey data to evaluate formal theories or to identify dynamic features of elections, for example, naturally found itself cramped by data gathered with different research questions in mind.

The general point here is that methodological concentration constricts the range of questions that political science (or any discipline) takes seriously. Conversely, methodological diversity would lead to a richer political science. Equipped with new methods, political science could take on new questions, because, for the first time, such questions *could* be pursued. Although experimentation is just one way to achieve the diversification that seems called for, its special advantages make it a particularly effective solution, as we begin to demonstrate next.

2. Experimentation Defined

In principle, experimentation may refer to a single form of scientific inquiry, but, in practice, experiments are nothing if not amazingly diverse. Experiments are undertaken in the laboratory and in the field. Experiments test the response of individuals, groups, neighborhoods, organizations, cities. Like experiments throughout the sciences, experiments in political science are drawn up with very different purposes in mind. Some are revelatory in aim, as in Milgram's (1974) famous (some would say notorious) demonstrations of obedience to authority. Others are carried out essentially for methodological and measurement purposes, a tradition inaugurated by Rice's (1929) experimental investigation of interviewer effects and sustained in the 1980s by a proliferation of experiments devoted to understanding the effects on the exression of public opinion due to question wording, format, and placement (see, for example, Schuman and Presser 1981; Tourangeau and Rasinski 1988). Still others are undertaken in the interest of learning about the effects of social policies, thereby heeding Donald Campbell's (1969b) plea for an "experimenting society." In their most celebrated and popular manifestation, experiments provide the empirical wherewithal for testing, refining, and even, on rare occasions, rejecting theory.

Just as experiments in political science serve multiple aims, they take up an astonishing variety of topics. Do ideologically charged campaigns induce ideological electorates? Does the complexity of the problems typically faced by public policymakers cripple their capacity to make sensible decisions? What circumstances—if any—can diminish the "tragedy of the commons": the failure of individuals to organize effectively for essential public goods? Is the work of congressional committees as chaotic as it seems, a subject fit only for journalism, or can it be interpreted in terms of general principles of agenda influence and group decision making? Is television news an "unseeing eye" that neither educates nor informs the American public, or does it shape the American public's conception of public life in powerful and pervasive ways?

All of this diversity is wondrous and, in many respects, admirable, but it does raise the question of what is it, precisely, that we are promoting. What *is* an experiment, anyway?

No matter how diverse experiments may be in practice, they share an interventionist spirit. Experiments *intrude* upon nature, and they do so (almost always) to provide answers to causal questions.[4] This is the distinguishing feature of Gosnell's experiment, conducted so many years ago. Gosnell had a causal proposition in mind—providing prospective voters with information about registration procedures will enhance the likelihood that they will make it to the polls come election day—and tested it by intervening in the natural, ongoing political process. When we undertake experiments, we are not in the business of "merely taking what comes," but, instead, we are "making observations in circumstances so arranged or interpreted that we have justification for analyzing out the factors relevant to our particular inquiry" (Kaplan 1964, 162).

It is the feature of intervention, and the control that such intervention brings, that distinguishes experimental research from other systematic empirical methods. In the fully realized experiment, the investigator seizes control

4. The notion of experiment is related in a close and complicated way to the notion of cause. For a much more detailed discussion of this point than we can afford here, and for an interesting if somewhat eccentric entree to the gigantic philosophical literature on cause, see Cook and Campbell 1979, chap. 1.

over the *production* of settings, the *creation* of treatments, and the *scheduling* of observations. The investigator does so in order to eliminate (or at least reduce) threats to valid inference. Settings are produced in order to exclude various nuisance factors that might otherwise interfere with the causal relation of interest. In this respect, the laboratory of the political scientist functions as the "lead shields, soundproof rooms, and sterile test tubes of the natural scientist" (Cook and Campbell 1979, 7). Treatments are created in order to isolate precisely the causal factor (or factors) of interest. In the ideal experiment, treatments differ from one another in systematic and specified ways, so that the effect due to a particular causal factor can be separated from the effects due to competing and naturally correlated factors. Finally, observations are scheduled in order to reduce the likelihood that the measured effects are contaminated by other causes. In these various generic ways, the experimenter intervenes in order to eliminate alternative rival interpretations, with the hope, not always realized, of being left with only a single, plausible interpretation.

In pursuit of interpretable comparisons, experiments in the social sciences these days characteristically feature both control groups (or multiple treatments) and random assignment. Multiple treatments may be created in an effort to minimize the effect of extraneous factors or to decompose a complex phenomenon or to test theoretically derived parametric predictions or to explore interaction effects between key variables. In all these cases, the purpose and advantage of the experiment is to manufacture precise and telling comparisons: "Good experimental design," as Campbell and Stanley (1963, 22) put it, "is the art of achieving interpretable comparisons." And, in the typical case, the experimenter can control which subjects—be they individuals, groups, organizations, cities, or whatever—receive which experimental treatments. Moreover, the experimenter typically exercises this control through *random assignment.* By randomly assigning subjects to treatments, the experimenter, in one elegant stroke, can be confident that any observed differences must be due to differences in the treatments themselves (within the limitations established by statistical analysis). By sweeping aside a host of alternative interpretations, random assignment is "the great 'ceteris paribus'" of causal inference" (Cook and Campbell 1979, 5).

Control groups and random assignment are unambiguously desirable features of experimental work in the social sciences; most of the experiments we present later and most of the experiments we admire possess both. But experiments need not include control groups and need not include random assignment. Consider, for example, Pascal's splendid experimental investigation of barometric pressure undertaken in the seventeenth century (recounted in Boring 1954, 577–78).

In 1648 the Torricellian vacuum was known to physics in general and to Pascal in particular. This is the vacuum formed at the upper closed end of a tube which has first been filled with mercury and then inverted with its lower open end in a dish of mercury. The column of mercury falls in the tube until it is about 30 in. high and remains there, leaving a vacuum above it. Pascal was of the opinion that the column is supported by the weight of the air that presses upon the mercury in the dish (he was right; the Torricellian tube is a barometer) and that the column should be shorter at higher altitudes where the weight of the atmosphere would be less. So he asked his brother-in-law, Perier, who was at Clermont, to perform for him the obvious experiment at the Puy-de-Dome, a mountain in the neighborhood about 3,000 feet ("500 fathoms") high as measured from the Convent at the bottom to the mountain's top. On Saturday, September 19th, 1648, Perier, with three friends of the Clermont clergy and three lawyers, two Torricellian tubes, two dishes and plenty of mercury, set out for the Puy-de-Dome. At the foot they stopped at the Convent, set up both tubes, found the height of the column in each to be 26 old French inches plus 3½ Paris lines (28.04 modern inches), left one tube set up at the Convent with Father Chastin to watch it so as to see whether it changed during the day, disassembled the other tube and carried it to the top of the mountain, 3,000 ft. above the Convent and 4,800 feet above sea level. There they set it up again and found to their excited pleasure that the height of the mercury column was only 23 French inches and 2 Paris lines (24.71 in.), much less that it was down below just as Pascal had hoped it would be. To make sure they took measurements in five places at the top, on one side and the other of the mountain, inside a shelter and outside, but the column heights were all the same. Then they came down, stopping on the way to take a measurement at an intermediate altitude, where the mercury column proved to be of intermediate height (26.65 in.). Back at the Convent, Father Chastin said that the other tube had not varied during the day, and then, setting up their second tube, the climbers found that it too again measured 26 in. 3½ lines.

Pascal lacked the physical means to create different treatment conditions, and he failed to randomly assign treatments to subjects, but his study is an experiment nonetheless. It reflects an interventionist spirit, ingeniously applied, and a sophisticated concern with ruling out alternative interpretations. Undaunted by an inability to create different pressures, Pascal opportunistically scheduled observations to take advantage of natural variations in pressure. Leaving a Torricellian vacuum at the base of the Puy-de-Dome provided some assurance that the changes the team discovered as they ascended the

mountain were due to changes in elevation and not to some more general change that momentarily swept through the region. It was wise to take several measurements at the summit under a variety of conditions, since if the weight of the air controls the height of the mercury column, then such variation should make no difference—and it did not. It was wise to take a measurement on the way down and highly reassuring to Pascal's claim that the mercury column showed an intermediate height. It was equally wise to demonstrate that the two tubes gave identical measurements at identical elevations, thus ruling out instrumentation differences as an explanation.

Naturally, Pascal's design would have been stronger had he had at his disposal many barometers, each randomly assigned to a different position on the mountain. It would have been stronger still had Pascal been able actually to *create* differences in barometric pressure, and to do so in such a way as to hold constant all other potentially relevant factors. Such a design—technically out of reach to Pascal at the time—would have neatly circumvented the problem that the observed changes in the mercury column might have been caused by factors correlated with elevation (such as differences in temperature). Still, Pascal's study beautifully illustrates that when we undertake experiments, we are not in the business of "merely taking what comes," but, instead, we are "making observations in circumstances so arranged or interpreted that we have justification for analyzing out the factors relevant to our particular inquiry." A preoccupation with control and a commitment to ruling out alternative interpretations, so clear in Pascal's study, are characteristic of good science in general and of good experimental science in particular.

Although the idea of control is central to scientific practice, and can be found prominently featured in the writings of Hume, Bacon, and especially Mill, the idea of control *group* is a relatively modern invention (Boring 1954). Control groups appeared first in biology in the later stages of the nineteenth century, most notably in a series of experimental investigations undertaken by Charles Darwin and Francis Darwin (Darwin 1875; Darwin and Darwin 1890). The use of control groups as recommended practice did not begin to creep into social science textbooks until the 1920s and 1930s and did not become standard practice in social science research until the middle of the century. Random assignment made its debut later still. Early users of control groups in biology and psychology ignored random assignment, as did Gosnell in his experimental study of turnout in 1924. It was perhaps the major contribution of Fisher's *The Logic of Experimental Design*, published in 1935, to insist on the importance of randomization to the integrity of experimental research. Fisher (1935, 19) put it this way:

> Whatever degree of care and experimental skill is expended in equalizing the conditions, other than the one under test, which are liable to affect the

result, this equalization must always be to a greater or lesser extent incomplete, and in many important practical cases will certainly be grossly defective.

Fisher's concern was that such inequalities not "impugn" the experimental comparisons; the remedy for such inequalities, of course, was randomizing treatments to subjects.

It is now common practice, if not taken entirely for granted, that experiments will include control groups and that subjects will be assigned to control and treatment groups on a random basis. Neither of these utterly familiar procedures was followed with any regularity just four decades ago in the social sciences, and neither has ever been standard practice outside the social sciences (except, perhaps, in some corners of biology). The emergence of control groups and random assignment is a reflection of the special problems faced by the social sciences. Each represents a means of coping with our common predicament: our inability to undertake research in perfectly controlled, closed systems; a preoccupation with dependent variables that are subject to a continuous onslaught of multiple causes; imprecise and implicit theories that typically do no better than muster ordinal predictions; and error-laden measurement. But these conditions do not always hold, and, when they do not, experimental research in the social sciences can resemble more the types of experimental research underway in the natural sciences (as we will see later in this volume).

For the moment, the main point is to avoid paralyzing squabbles over exactly what is, and exactly what is not, an experiment. Experiments are preoccupied with control and with ruling out alternative interpretations, and, to these ends, control groups and random assignment are often, but not always, essential. Along with Campbell and Ross (1970, 123), we maintain that methodological prescriptions, standards for acceptable practice, do not emerge "as logical dispensations from the philosophy of science or mathematical statistics," but from the "iteration of effort and criticism," the thick, rich experience of actually doing social science.

3. Experimental Strengths

We have already noted the general benefits that a commitment to experimentation would bring to the study of politics. Experimental studies, taken seriously and deployed widely, would provide much-needed diversification and would open up lines of investigation unimaginable under prevailing methods. Experimentation is also attractive because it delivers particular advantages largely unavailable through other means. Compared to other social science methods, experimentation provides a clearer glimpse of cause-and-effect relations, enables complex phenomena to be decomposed, accelerates interdisciplinary conversations, especially with psychology and economics, is more likely to produce anomalous facts that must be taken seriously, and travels more flexibly across different levels of aggregation. We develop each of these comparative advantages next.

Testing Causal Propositions

No doubt, experimentation's most emphasized advantage is its capacity to test cause-and-effect relations. This virtue of experimentation is vividly displayed in Cook and Campbell's introduction to the idea of experiment.

> The word *experiment* denotes a test, as when one experiments with getting up two hours earlier to see if this makes one's working day more productive. The test is usually of a causal proposition: for example, does garlic or curry add a better flavor to certain rice dishes? There are some uses of the concept of experiment where the link with cause is not immediately obvious, yet still paramount. For instance, an airplane is "experimental" only if one wants to test whether it flies faster, more efficiently, or more safely than some alternative.
>
> The notion of a "trial" or deliberate manipulation is also linked to experimenting. Actually getting up earlier on some mornings is the most direct way of evaluating how one's productivity changes; using curry on some occasions and garlic at others will enable one to evaluate which seasoning improves the rich dish; and without flying the experimental airplane, it will be difficult to test. (Cook and Campbell 1979, 2–3)

The unrivaled capacity of experiments to provide decisive tests of causal propositions follows immediately from two aspects of control emphasized in experimental practice. By creating the treatments of interest, the experimenter holds extraneous factors constant and ensures that subjects encounter treatments that differ only in designated ways. By assigning subjects to treatments randomly, the experimenter can be confident (within the limitations established by statistical inference) that any differences observed between subjects assigned to different treatment conditions *must* be caused by differences in the treatments themselves. Defined in this way, experimentation can become a tool of unexcelled power and precision, one that, for some purposes, has no serious competition.[5]

5. Experimentation does not eliminate *all* threats to internal validity. Those not eliminated include some statistical pitfalls and several problems that are typically associated with large-scale policy experiments conducted in the field (Cook and Campbell 1979).

One terrific example of an experimental study providing a crisp causal answer is supplied by research on the ancient questions of how—and how well—ordinary citizens come to their views of public life. Much of the considerable empirical effort devoted to these questions in the last quarter-century has focused on the specific possibility that Americans might approach the political world equipped with ideological frameworks. But to analysts of national surveys undertaken in the 1950s, most Americans seemed quite innocent of ideology. According to Converse (1964), who pushed this claim hardest, ordinary Americans were bewildered by ideological concepts, had no consistent point of view on government policy, and possessed authentic opinions on only a handful of pressing questions.

In the last decade, however, various revisionists have argued that Converse's conclusions no longer hold. Ideological innocence reflects not the incorrigible limitations of ordinary citizens, but, rather, the comparative tranquility and intellectual blandness of the Eisenhower years. Furnish Americans with an ideological politics, so argue the revisionists, and they are quite capable of responding in kind.

The empirical centerpiece of the revisionist argument was the compelling demonstration (so it seemed at the time) of greater cohesion in the American public's beliefs on policy beginning in the 1960s. In *The Changing American Voter* (1979), Nie, Verba, and Petrocik replicated Converse's original analysis, this time making use of a series of national election surveys that spanned more than two decades. Through the 1950s and early 1960s, Nie and associates found correlations in the general public between opinions on a variety of topical issues to be puny—just as Converse had. This changed dramatically in 1964, as the ideologically charged Johnson-Goldwater presidential campaign reached its natural climax. Suddenly, public opinion became more cohesive: opinion on school integration became aligned with feelings about the size of government, views on foreign policy became linked to beliefs regarding the government's responsibility to subsidize health care, and so on. Detected first in 1964, this newfound ideological structure to public opinion persisted at about the same level through 1968 and has since slightly diminished. Nie, Verba, and Petrocik interpreted these results to signal a sea change in public thinking, one set in motion by the ideologically tumultuous campaign of 1964.

Perhaps. But, also in 1964, the national election study tinkered with the survey questions used by Converse and his revisionist critics as prime evidence in the debate over ideology. The changes appear innocuous enough: from a format in which respondents were asked to declare how much they agree or disagree with a certain policy to an arrangement in which respondents were asked to choose between a pair of opposing alternatives; and from a gentle to a somewhat more insistent invitation to confess to no opinion one

way or the other. These alterations seem minor, but, because they were introduced at the precise moment of the apparent dramatic change in American public opinion, and because we know from other research, much of it experimentally based (e.g., Schuman and Presser 1981), that ostensibly small changes in question wording can sometimes produce sizable differences in opinion, such tinkering constitutes a rival explanation of some plausibility. In Campbell and Stanley's (1966) terminology, the observed change in public opinion between 1960 and 1964 might be due to change in *instrumentation*, nothing more. How can we tell whether the toning up of public opinion detected in 1964 was caused by a transformation in the character of American electoral politics or merely by a change in the way public opinion was assessed?

We could undertake an experiment. Or, better, we could read the report of Sullivan, Piereson, and Marcus's splendid experiment, originally published in 1978 and reprinted here. Residents of the Twin Cities region of Minnesota were interviewed regarding their opinions on a variety of political issues. Half of the respondents (by random determination) were asked for their views using the pre-1964 question format; the other half of the sample was questioned using the question format introduced in 1964. In this elegant experimental design, the character of American electoral politics is obviously held constant—both groups are interviewed at the same time in the same political setting—while the opinion assessment technique is systematically manipulated.

It turns out that this experimental manipulation produced large differences in the pattern of relationships between opinions on government policy. Moreover, and this is the decisive point, these differences mimic, in a remarkably fine-grained way, the differences that Nie and associates had attributed to transformations in the nature of American politics. In light of these experimental results, most if not all of the change in the structure of public opinion witnessed in 1964 now appears to be artificial, induced not by basic political alterations but by mundane modifications in question wording—a pure if horrifying example of instrumentation change masquerading as real change. This conclusion, of course, radically inverts a major claim of *The Changing American Voter.* Boisterous events and panoramic changes did indeed mark American politics in the 1960s; in the meantime, however, American public opinion on government policy appeared quite undisturbed by all the commotion.

A second example that illustrates the experimental method's comparative advantage in detecting cause-and-effect relations is provided by recent research on the political impact of television. For their news about politics, Americans must depend upon information and analysis supplied by others—in modern times, upon information and analysis supplied by mass media, especially television. This line of argument was introduced by Lippmann (1922 and 1925) and elaborated upon by others since (especially Cohen 1963; Converse 1975; Downs 1957; Lazarsfeld and Merton 1948). But how should the claim be tested?

One common approach goes like this. Suppose we are interested in the influence television news might exercise over the American public. Like many before us, we therefore decide to interview a sample of Americans, carefully selected to be representative of the nation as a whole (random selection of the sample, not random assignment into experimental and control conditions). We partition our sample into two groups, those who tell us that they rely primarily upon television news for their information about politics and those who say they rely on other sources. When we compare the political views of the two groups, we discover that the television-reliant group more often regards unemployment as the country's most serious problem and more often evaluates the president's performance primarily on how well, as they see it, he has managed the national economy. Meanwhile, we undertake a content analysis of television news coverage, finding that just before and during the period of interviewing, television news has been unusually preoccupied with unemployment. We conclude that television news does, indeed, shape its viewers' conceptions of political reality.

Although there is nothing silly about this hypothetical study, it does have serious limitations, just those that a true experiment counteracts. First and foremost, it cannot establish causal relationships. Observing that television news coverage and viewers' beliefs are correlated is not the same as establishing that television coverage influences viewers' beliefs. Citizens who rely primarily on television may differ in many ways from those who obtain their information elsewhere, and it may be these differences that are responsible for generating different outlooks on national problems. If the television group is disproportionately working class, for example, their special concern about unemployment might be due not to television coverage but to their own experiences in the labor force, or the experiences of their friends and co-workers. Although we might be able to test this particular explanation by partitioning the sample into different occupational groups and then seeing whether the same relationship between television viewing and political outlook was maintained within each occupational group, we could never know for certain whether *all* such rival explanations had been ruled out.

This is debilitating for causal inference, and it is exactly the problem overcome in experimentation through random assignment. Randomly assigning some people to television and others to newspapers (ignoring for the moment any ethical and practical difficulties that might stand in the way), the experimenter can be certain that whatever differences detected between the two groups can be traced to differences in the treatments. Alternative explanations due to preexposure differences—associated with class, unemployment experiences, *or anything* else—are dispatched by this simple procedure.

This virtue of random assignment is illustrated in a series of recently completed experiments on television news (reported in full in Iyengar and Kinder 1987; a report based on a subset of the experiments appears here in section 4). In these experiments, systematic alterations were unobtrusively introduced into the television news broadcasts viewed by ordinary citizens. Thus, citizens randomly assigned to different conditions were furnished with slightly different glimpses of the political world. Taken as a whole, these experiments support two general conclusions: that television news powerfully influences which problems viewers regard as the nation's most serious (agenda setting); and that, by highlighting certain aspects of national life while ignoring others, television news sets the terms by which political judgments are rendered and political choices made (priming). Are these causal conclusions, based on a series of experiments, infallible? Well, no. We will return to these experiments, and their particular limitations later, when we take up experimental shortcomings as a general subject.

Analytic Decomposition

By creating treatment and control conditions, the experimenter is able to isolate one causal variable at a time. This, in turn, allows complex phenomena to be decomposed in a way that is impossible under more passive research strategies. Experimenters need not wait for natural processes to provide crucial tests and telling comparisons: they can create them on their own. In this respect, experimental control confers a whopping comparative advantage.

Two of the chapters reprinted in Part 3 of this book illustrate the power of experiments to isolate the effects of different causal variables. Both examine individual behavior and the resulting group outcomes that arise in the decentralized provision of collective goods. The central difficulty that must be overcome in the generation of such goods is the inclination among individuals to free ride: to benefit from the collective good without bearing the cost of its provision.

The experiment reported by Isaac, Walker, and Thomas, first published in 1984, focuses on the identification of factors that might affect the extent to which free riding occurs. Their design isolated two causal variables in particular: the private gain to be realized from free riding and the size of the group. In naturally occurring groups, these two factors are difficult to disentangle, and the first is hard to measure. It has long been theorized by political scientists, economists, and sociologists that larger groups will fall prey to greater levels of free riding (Chamberlin 1978; Marwell and Ames 1980 and 1981; Olsen 1965; Palfrey and Rosenthal 1984 and 1988; Roberts 1976). But Isaac, Walker, and Thomas show that when private incentives to free ride are controlled, the effect of group size *by itself* is, if anything, just the reverse: as group size increases, free riding declines. It is difficult to imagine being able to provide such convincing empirical evidence on this important question other than by experimentation.

In the second experiment, Ferejohn, Forsythe, Noll, and Palfrey consider the free-rider problem when it is conjoined with a "coordination problem." Not only must a group decide how much of a collective good to provide and who should bear the cost, but *which* of several public goods to produce. The different projects cost different amounts, individuals value them differently, and budget constraints preclude the possibility of doing them all. Of course, there is also the free-rider problem that each member would rather have the rest of the group bear the costs. Ferejohn and company's experiment examined four factors that might affect the ability of the group both to overcome the free-riding problem and to solve the coordination problem: experience with such problems, heterogeneity among group members, financing details, and voting rules. In naturally constituted groups, these factors tend to be confounded, making it virtually impossible to disentangle their independent effects. Not so in experimentally constituted groups, where the creation of conditions can "unconfound" and decompose the relevant causal factors.

As a final example, consider again the relationship between television news and public opinion. Experimentation is useful not only in helping to establish the claim that television news influences public opinion, but it also opens the way to investigating the particular aspects of television coverage that are responsible for that influence. For example, with experiments, it is possible to test pointedly whether the news stories that lead off the broadcast are more influential merely because of their privileged position (they are); whether the dramatic personal vignettes that the networks commonly use to exemplify national problems, which make for riveting television, are especially influential (they are not); or whether the magnitude of priming depends upon how deeply the news implicates the government (it does; all these results are reported in detail in Iyengar and Kinder 1987).

Interdisciplinary Ties

As political science has become methodologically specialized, it has become, to a degree, methodologically isolated from its cognate social sciences. The development of specially tailored methods within political science may have had the consequence of slowing down communication across disciplinary boundaries. Of course, exchange across disciplinary boundaries is always slow. Too often the borrowing discipline is touting with great enthusiasm a theory or technique that the donor discipline is about to abandon as obsolete.

Communication across disciplinary lines, which can be catalytic to significant advances in knowledge, would be accelerated if political science were more diversified methodologically.

Experimentation is especially attractive in this regard. For whatever reason, the recent emergence of experimentation as a bona fide approach to empirical research in political science has simultaneously served as an inspiration for, and been a by-product of, interdisciplinary collaboration.[6] That political science would benefit from an exposure to alternative social science perspectives is hardly controversial. But successfully bringing these different points of view to bear in a productive fashion is another matter. Why is it that the experimental approach has been so successful at overcoming traditional disciplinary hurdles?

One explanation is historical. Laboratory experiments in the social sciences appeared first in other disciplines, such as psychology and economics, so that the principals involved with experiments early on in political science were often trained in other disciplines and naturally interested in problems at the interface of their original discipline and political science. In this regard, it should come as no surprise that experimental political science and interdisciplinary research go hand in hand, if only as a matter of historical accident.

But there is more than history at work here. Different approaches, particularly different *empirical* approaches, entail posing different questions about essentially the same phenomena. In the experimental approach, investigators go to great lengths to identify and isolate the effects of specific variables, holding everything else as constant as possible. Consequently, it is quite natural that the theoretical models used in conjunction with experimental data are ones that systematically and rigorously formalize the effects of one set of specific variables on other specific variables. Furthermore, theories that are highly parametric and formalized make predictions that are often too detailed and precise to hope to obtain anything close to an appropriate test using data that are not experimental, either because certain key variables are unmeasurable or because there are simply too many conditions varying simultaneously. This produces a natural synergy between formal, positive theories of politics, imported largely from economics, and a brand of experimental investigation that also has its origins in economics. This has produced new ways to study political processes from both the theoretical and the empirical side.

In a similar war, the importation of experimental methods from psychology, especially to study individual information processing and decision making, has had a significant impact on how we think about public opinion, voters and elections, and other political phenomena. We ask new questions, again partly because the use of experimental methods offers us a way to answer

6. As illustrated by this essay and volume.

them accurately and partly because the data from experiments suggest questions that no one might have thought of asking otherwise (e.g., Tversky and Kahneman 1981). Indeed, the very meaning of opinion is currently under reassessment, provoked by new findings and theories in cognitive psychology (Kinder and Sanders 1990; Tourangeau and Rasinski 1988; Zaller 1985). The interdisciplinary spirit of experimental political science is unquestionably helping to speed up the infusion of knowledge from other sectors of the social sciences.

A recent instance of useful interdisciplinary exchange between economics and political science is the incorporation of rational expectations into models of dynamic processes. An obvious and often emphasized feature of such processes is that individual behavior and outcomes are heavily driven by expectations. Therefore, models of dynamic processes have little predictive or explanatory power unless they impose assumptions about these expectations that, in turn, imply testable restrictions on observable data. Rational expectations theory generates one set of restrictions that derive from equilibrium assumptions about the accuracy and rationality of expectations. The implications of rational expectations for dynamic aggregate behavior can be traced back to the 1950s in both political science and economics (Grunberg and Modigliani 1954; Muth 1961; Simon 1954). Political scientists were interested in the evolution of public opinion and the information content of poll results, while economists were interested in the evolution of the macroeconomy and the information content of prices. Models incorporating rational expectations were quick to catch on in economics, but—remarkably—political science paused for thirty years, until the pathbreaking work of McKelvey and Ordeshook (1984 and 1985; Collier et al. 1987) on the effect of poll results on voter beliefs and electoral outcomes. The McKelvey and Ordeshook reading in Part 4 of this volume evolved from their earlier experiments that examined the informational effects of interest-group endorsements and poll results. Even more recently, other dynamic models borrowed from economics have provided the theoretical background for additional laboratory elections (Boylan et al. 1991).

This can be a two-way street, of course. Coincident with the resurgence of interest in using principles of rational forecasting in dynamic formal models in political science (e.g., Austen-Smith and Banks 1989; Baron and Ferejohn 1989; Cukierman 1985; Ledyard 1989; Ordeshook and Palfrey 1988), has been the incorporation of explicitly political factors in state-of-the-art macroeconomic theory. Alesina (1987), Kydland and Prescott (1977), Nordhaus (1975), Rogoff and Sibert (1988), and others have examined the interaction between elections and macroeconomic policy in an attempt to understand such phenomena as the political business cycle. This line of work points to the possibility of experimental research investigating the effects of political institutions on macroeconomic performance (see, e.g., Boylan et al. 1991).

For sheer interdisciplinary breadth, it is hard to beat the study of collective action (Dawes 1980; Olson 1965; Samuelson 1954). Most behavioral sciences are well represented. Psychologists and sociologists are interested in studying how and why individuals behave as they do when they are confronted with a decision problem that pits their personal desires directly against some notion of social good. How can such behavior be explained as a function of personal characteristics of the individuals (e.g., income) and characteristics of the group (e.g., close knit vs. anonymous)? Political scientists add institutional features to this list of explanatory variables, focusing on problems of specific political relevance. By and large, economists look for general incentive effects describable in terms of theories of individual behavior that include only the "economically relevant" personal characteristics (such as preferences and costs) as independent variables, meanwhile viewing the institutional features as part of a design problem, much as an engineer would create and (experimentally) test theories about how to select materials to build a road. Together, these diverse perspectives cover a large patch of the interesting intellectual territory in empirical research on collective action (and are represented in the readings that follow in Part 3).

Perhaps this sounds too good to be true. Several disciplines attacking a multifaceted research problem from different angles certainly has a utopian ring to it. Such an arrangement works in practice only when there is communication between the researchers and, at least to a degree, there is some standardization of methodology. It is fair to say that, at least on the empirical side, experimentation has been the primary focusing device in this instance. Many of the collaborations cross disciplinary lines (Dawes and Orbell; Ferejohn, Forsythe, Noll and Palfrey, to name two). Perhaps more important, most researchers are aware of the work being done in other disciplines on this problem, cite each other's work, use each other's work to develop alternative testable theories and new experimental designs, and even meet together at multidisciplinary conferences. In short, experimental approaches to empirical political science seem especially likely to generate interdisciplinary interaction, particularly for problems in which there is a natural meeting between the study of institutions (political science), incentives (economics), and individual and group decision making (psychology and sociology).

"Stubborn Facts" and Theoretical Invention

Empirical results are often the parent of theoretical invention. When results challenge orthodox understandings, and when they cannot be dismissed, they can lead to genuine advances. Following Cook and Campbell, we

. . . find much to value in the laboratory scientist's belief in "stubborn facts" that "speak for themselves" and which have a firm dependability greater than the fluctuating theories with which one tries to explain them. Modern theorists of science—Popper, Hanson, Polanyi, Kuhn, and Feyerabend included—have exaggerated the role of comprehensive theory in scientific advance and have made experimental evidence almost irrelevant. Instead, exploratory experimentation unguided by formal theory, and unexpected experimental discoveries tangential to whatever theory motivated the research, have repeatedly been the source of great scientific advances, providing the stubborn, dependable, replicable puzzles that have justified theoretical efforts at solution. (Cook and Campbell 1979, 24)

It is no accident that Cook and Campbell refer explicitly to *experimental* discoveries. When experimental discoveries produce anomalies, they are less apt to be dismissed. It is impossible to replay history and expensive to redo public opinion surveys. But experiments can be replicated. Moreover, by custom, experimental procedures are described in sufficient detail so that replication is practical. And, finally, experimentation breeds an ethic of replication, so that experiments, again by custom, are much more likely to be replicated. The experimental discovery of anomalous results that survive replication is likely to be taken seriously, and may lead in time to better theory.

A conspicuous example of replicated anomalies of high relevance to political science can be found in the experimental work of Kahneman and Tversky. In a series of ingenious experiments, Kahneman, Tversky and others have uncovered a catalog of systematic departures from "rational" decision making (Tversky and Kahneman 1981). Among other things, they have demonstrated that sizable shifts in choice can be produced by "seemingly inconsequential changes in the formulation of choice problems" (1981, 453): framing the choice in one way rather than in a logically equivalent way can radically alter which option is chosen and which options are foregone. Such discoveries, which are anomalous from the perspective of orthodox rational choice theory, have had a sensational impact on theory and research on decision making throughout the social sciences. The empirical literature is massive (for partial reviews, consult Abelson and Levi 1985; Kahneman, Slovic, and Tversky 1982; Machina 1983), and because the basic results have proven robust (Grether 1980; Grether and Plott 1979), a good bit of theoretical invention has followed (e.g., Chew 1983; Dekel 1986; Kahneman and Tversky 1979; Machina 1982; Thaler 1980). Although the original work on heuristics and biases in individual choice began naturally within psychology, it is now prominently featured in studies of political judgment and decision making: in Iyengar and Kinder's (1987) experimental research on television

news and priming; in Quattrone and Tversky's (1988) sketch of a psychological model of voter decision making (reprinted in Part 2); in Jervis's (1976) account of foreign policy decision makers wrestling with great complexity and uncertainty; and more.

As we noted earlier, virtually all of the departures from rationality have been replicated extensively, though entirely within individual choice environments. Of great interest to political scientists and economists, naturally, is whether the individual biases persist in the aggregate. Obviously a key question here is whether aggregation serves to accentuate or diminish individual departures from rationality. Do many small errors by different individuals tend to cancel each other out or do they cumulate to an avalanche of irrationality?[7] We do not know the answer yet, but it is typical of the kind of provocative question that can emerge from simple, carefully constituted, and easily replicated laboratory experiments that, by design or chance, produce the stubborn facts that can challenge theoretical orthodoxy.

Another example that may, in time, prove as provocative as the Kahneman and Tversky discoveries are the election experiments of McKelvey and Ordeshook (1986 and Part 4). In a series of experiments, McKelvey and Ordeshook have consistently found evidence suggesting that candidates will be responsive to voters even when voters are miserably informed about politics. Their results reinforce the idea that retrospective voting by information-poor voters is effective at disciplining candidates. At the same time, the findings cast doubt on widely held views that democracy depends upon a reasonably high level of political awareness among the voters. In fact, if the laboratory elections conducted by McKelvey and Ordeshook are indicative of candidate and voter behavior in mass elections, then voters need not know anything specific about the issues of the election or the positions taken by candidates. The same outcome results when each voter decides whether to support the incumbent based solely on how well off that voter was under the current administration compared to past administrations. These are fascinating and provocative results. How "stubborn" they are, and how great an effect they ultimately will have on our understanding of elections, it is too early to say. At the least, McKelvey and Ordeshook's method and results have brought into clear focus a fundamental question for democratic practice.

Flexibility Across Levels of Aggregation

A somewhat different dimension of the diversity of the experimental method is its wide range of application across different levels of aggregation. Few

7. Knez and Smith (1987), Samuelson and Bazerman (1985), Kagel, Harstad, and Levin (1987) and others have pursued similar aggregation questions in economic environments.

would dispute that it is desirable for an empirical method to have a firm footing at the individual level, yet lend itself to empirical analysis across a broad spectrum. This is especially so in political science, where relevant applications exist at all levels of aggregation, from the study of behavior of large groups of individuals in elections to bilateral negotiations between two superpowers. Experiments have the flexibility to test theories and to provide empirical insights at all levels.

Nowhere is this juxtaposition of the individual and the aggregate more evident than in the study of representation. On the one hand, the basic foundation for the legitimacy of authority in democratic systems is that outcomes are the representative product of a diverse collection of separate, autonomous, individual choices. How do individual citizens decide whether or not to vote? How do they decide which candidate to support? How do they form opinions about candidates and issues? What lessons do citizens draw from the media about pressing national problems? On the other hand, that which is the ultimate object of interest to many political scientists happens at a highly aggregated level and falls under that impressive and almost ethereal heading of "Mass Behavior." Does high turnout favor Democrats or Republicans? What is the relationship between campaign spending and incumbent success? Is momentum a real force in presidential primary elections or merely a metaphor? Do the media educate or mislead the public regarding the nation's condition? Of course, these two extreme levels of aggregation have important links. Knowledge about individual voting behavior will certainly shed light on the effects of high (or low) turnout. And between these two extremes—the individual citizen and the mass public—are many intermediate levels of analysis: legislatures and committees, juries, courts, interest groups, and so on. While these examples are taken mostly from the study of American politics, similar applications can be found in international relations, comparative politics, and other fields of political science.

To illustrate the point we are trying to make about experimental flexibility and levels of aggregation, consider the following example. A fundamental fact to emerge from research on public opinion over the last four decades is that Americans know astonishingly little about the world of politics. Large portions of the American public do not know whether it is the Contras who are Communists, how William Rehnquist makes a living, who exactly represents them in the U.S. Senate, and on and on (Kinder and Sears 1985). On matters of information, there exist staggering and astronomical differences between those in leadership positions, who command "vast treasures of well organized information," and most of the rest of us, whose grasp of political events is partial and precarious. "Very little information 'trickles down' very far" (Converse 1964, 212). The big aggregate question, of course, and one that surprisingly few have asked, is: Does the quality of policy outcomes in democratic societies depend in any significant way on the level of information of the typical voter? Does lack of information at the individual level translate into poor outcomes at the aggregate level?

Rudimentary analysis and common sense suggest that a well-informed electorate is necessary for a democracy to function well (e.g., Thompson 1970). Some even posit a necessary relationship, feeling neither the logical obligation of proof nor the empirical obligation of evidence. Others are acutely aware that the degree of the relationship depends, in some difficult to specify way, on what information voters might want to use and how they might use it. Key (1966), for example, suggests that for an electoral system to perform effectively, voters may only need to use a retrospective test that requires the simplest of information, available to even the least sophisticated voters (also see Fiorina 1981). Still others have suggested that aggregation itself cancels out individual ignorance and eccentricity, so that the *public* behaves sensibly and reasonably, as if it were constituted by the well-informed and thoughtful citizens that successful democracy would seem to require (Converse 1990; Page and Shapiro 1992). This is a big and complicated question, to which experiments can usefully contribute. As we noted earlier, at the aggregate level, the McKelvey and Ordeshook results indicate that perhaps we should rethink conventional wisdom about the importance of an informed electorate for the success of a democracy. Poor information at the individual level may not have as serious policy consequences as many democratic theorists claim.[8]

More generally, experimental data can contribute to our empirical understanding of political phenomena at several different levels of aggregation. Another example, at various intermediate levels of aggregation, is the examination of institutional rules for committee and legislative behavior, discussed in the introduction to Part 5. And for another, many of the "collective action" experiments of the sort included in Part 3 are *explicitly* concerned with studying how behavior changes as groups get larger. In these experiments, the level of aggregation is itself a manipulated variable.

Finally, we wish to deal preemptively with a particular skepticism we have encountered regarding the experimental study of mass behavior. What can we learn about *the public* from experimental studies of, at most, 100 individuals? A lot, we think. In the first place, much can be learned about mass behavior by understanding individual behavior. Thus, for example, the public opinion experiments discussed earlier are probing theories about individual behavior that have immediate implications for aggregate political phenomena, such as momentum, bandwagon effects, and shifts in aggregate

8. The aggregate implications of *biased* information, as opposed to simply poor information, is another matter. We know of no experiments of this sort to date.

opinion. Second, there is a fair amount of evidence from economics suggesting that theories about mass behavior that were long thought to apply only to very large groups, say the empirical validity of competitive market equilibrium, are, in fact, quite robust to the size of the group (see, e.g., Smith 1982). How large a group is needed to produce "mass" behavior may be surprisingly small in many instances. Third, while few experiments have been run with groups larger than 50, experimenting in a laboratory with very large numbers of subjects can certainly be done in principle, and perhaps this is an appropriate place to call for more of it. Presumably one reason why such experiments are rarely carried out is because of their expense. But compared to the costs of gathering, storing, distributing, and administering large data sets (as, for example, the Bureau of Labor Statistics does), or of operating a large-scale survey research organization, or of conducting experiments in physics, chemistry, and biology, the cost of conducting both laboratory and field experiments in political science is remarkably reasonable. Finally, lest we forget Harold Gosnell, mass experiments not only can be done but have been done: Gosnell's experiment on turnout was based on 6,000 adults scattered across 12 Chicago districts. Political scientists should not be afraid, at least on occasion, to think big when it comes to experimentation.

4. Experimental Shortcomings

Like other methods, experiments have liabilities as well as strengths. We acknowledge and briefly review them here under three headings: the inevitable ambiguity surrounding the meaning of experimental treatments, the unsuitability of experimental methods to some core problems in political science, and some hazards that inescapably attend experimental generalization. Of the three, we pay primary attention to the problem of generalization, both because it occupies such a prominent place in arguments against experimentation in political science, and because such arguments seldom rise much above sneering references to college sophomores, on whose backs most experiments run. Our general message is that the shortcomings of experimentation, while they are real and cannot be avoided entirely, do not constitute anything like insuperable obstacles. Indeed, the creative and energetic experimenter can even, on occasion, convert these apparent liabilities into strengths.

Ambiguity of Experimental Treatments

The problem of ambiguity pertains to the interpretation of experimental effects. In an ideal experiment, we put ourselves in position to conclude that the differences we observe in behavior across the experimental treatments are caused by differences in the treatments themselves. But while experimental

methods enable us to draw causal inferences of exactly this form, they do nothing to reveal to us the meaning of the experimental treatments that we create. We can say, with special authority, that the treatment did it, but what, exactly, is the treatment?

Carlsmith, Ellsworth, and Aronson (1976) call this the problem of "multiple meaning" and refer to it as "one of the most perplexing and pervasive problems in experimentation" (61). Understanding experimental treatments and, therefore, experimental results involves an assessment of the correspondence between conceptual variables—such as "impoverished information" or "agenda control" or "television news"—and their empirical realizations. No rules prescribe exactly how to move between these two levels. Because political concepts are ambiguous and experimental operations unstandardized, there is no necessary one-to-one correspondence between empirical realizations, on the one hand, and their conceptual referents, on the other.

From a certain point of view, this diversity is actually quite appealing. Insofar as different empirical realizations show convergent results, we can be more confident that the results are rightfully interpreted in terms of the intended conceptual referents. Thus, one powerful response to the problem of the ambiguity of experimental treatments is to undertake systematic replications. When very different operational realizations of the experimental treatment produce essentially the same result, the meaning of the experimental treatment becomes less ambiguous.

Consider this example. In 1959, Aronson and Mills published an experiment that seemed to demonstrate, consistent with cognitive dissonance theory, that when people undergo a severe initiation in order to be admitted to a group, they will, as a consequence, find the group more attractive. In this experiment, the conceptual variable, severity of initiation, was empirically realized by embarrassing some subjects and not others. The severe initiation treatment consisted of having the college women subjects read aloud (in the presence of a male experimenter) a list of obscene words and two vivid accounts of sexual activity taken from contemporary novels. But is embarrassment a proper realization of the severity of initiation? And is embarrassment actually responsible for greater affection for the group, or is it something else? Aronson and Mills's claim is considerably strengthened by a systematic replication conducted by Gerard and Mathewson (1966), who found the same result when manipulating the severity of initiation in a very different way. Instead of embarrassing their (severe initiation) subjects by having them read sexually explicit passages aloud, Gerard and Mathewson administered mild but unpleasant electric shocks to theirs. These experiments may be grisly in their details, but they are exemplary in their display of the power of systematic replication to pin down the meaning of experimental treatments.

A related example is provided by Isaac, Walker, and Thomas's (1984) experiment on the effect of group size on the provision of a public good. Previous research had defined group size in a way that had confounded it with individuals' marginal incentives to participate. Thus, what seemed, in past work, to be the effect of group size had actually been the effect of an amalgam of group size and individual preferences. Recognizing this, Isaac, Walker, and Thomas manipulated each independently and found, contrary both to previous results and to the conventional wisdom, a positive effect of group size on individuals' willingness to contribute to the group effort. In this case, a systematic replication clarified the ambiguous result left behind by previous experiments.[9]

Limitations in the Scope of Experimental Methods

Some problems that are central to political science simply cannot be investigated by experimental means. If we happen to be interested in understanding every historical detail of the Reagan landslide of 1984 or the shaping of Southern politics at the turn of the century or the relationship between class and participation in public life, then we must rely on methods other than experimental. For the most part, history and social structure lie beyond the reach of experimental manipulation. But to admit that experimentation cannot address all problems that the field deems worthy of investigation is not a special confession: all methods suffer such limitations. There are enough problems left, after experimentation's imperfect reach is acknowledged, where experimental methods can make a special contribution. Moreover, it seems to us that the limitations in the scope of experimental methods are often exaggerated. Historical moments can be recreated and studied in experiments, as in Iyengar and Kinder's (1987) restaging of the final days of the 1980 presidential campaign; large, complex, and cumbersome political processes can, when stripped to their essentials, be brought under experimental control, as in McKelvey and Ordeshook's laboratory investigation of elections (1986 and Part 4). Such studies suggest a wider scope to experimental methods than is commonly appreciated. They obviously also raise, and in a particularly acute form, the question of the generalizability of experimental results, which is our next topic.

Generalization

While thoughtfully designed and carefully executed experiments are uniquely strong on matters of *internal* validity—providing evidence on causal

9. For more and more detailed recommendations about how to cope with the problem of the ambiguity of experimental treatments, see Carlsmith, Ellsworth, and Aronson 1976 and Cook and Campbell 1979.

relationships—they are often criticized on matters of *external* validity—providing assurance about the generalizability of results. Naturally, it is generalizable results that we are after. Experimental findings are of interest to most political scientists insofar as they bear on the workings of real political processes. But to generalize from particular experimental arrangements and populations to the real political world is to participate in what Campbell (1969a) has called "the scandal of induction." Always and inescapably, generalizations are matters of opinion.

Concern about the generalizability of experimental results in particular usually takes one of three forms. First, because experimental participants ordinarily know that they are taking part in the study of something (even if they are not sure what), this knowledge alone may induce alterations in their behavior. This concern arises equally in experiments with individuals, groups, organizations, companies, school systems, cities, and all the rest. In all such cases, experimental participants might become more attentive than they otherwise would be, knowing that they are the object of study, or they might become less. They might defer to the experimenter's authority or they might react against it. Each response represents a reaction to the special and, in some respects, *artificial* nature of the research setting. Second, experiments are often conducted with samples of convenience, leading to skepticism over whether experimental results can be generalized safely to the populations of real interest. For American social scientists situated in universities, no population is, of course, more convenient than the local student body. And the typical college sophomore, as Hovland (1959) warned some years ago and as Sears (1986) has recently documented, may be a rather peculiar creature. Third, experimental results are always subject to the charge that they depend precariously on exactly how the independent variables were created. Results that emerge in one experiment might disappear in another, when conditions are realized in a slightly different way; or a finding that seems startling and important sometimes turns out to be the product of an unrealistically powerful manipulation, one that rarely occurs in natural settings.

These concerns about generalizing from experimental results—be they generalizations across settings, populations, or treatments—always accompany the presentation of experimental results, and so they should. This acknowledgment may seem uncontroversial, but it does run against official pronouncements and customary practice in psychology, where such concerns tend to be brushed aside. For example, some years ago, Campbell (1969a, 361) recommended that psychology should follow the "successful sciences" and not worry so much about representative sampling, generalization, and the like.

> . . . Typical of science is the case of Nicholson and Carlisle. Taking in May, 1800, a very parochial and idiochronic sample of Soho water,

inserting it into a very biased sample of copper wire, into which flowed a very local electrical current, they obtained hydrogen gas at one electrode, oxygen at the other, and uninhibitedly generalized to all the water in the world for all eternity. . . . In the successful sciences, generalizations have never been "inductive" in the sense of summarizing what had been observed within the bounds of the generalization, but instead have always been presumptive, albeit guided by prior laws. The limitations to the generalization have emerged from checking in nonrepresentative ways on initial bold generalizations. Scientists assumed that hydrolysis held true universally until it was shown otherwise.[10]

Such recommendations both influence and reflect actual practice. As Sears (1986) and others have shown convincingly, American social psychology has come to rely overwhelmingly on college students undertaking academiclike tasks in laboratory settings. The new and rapidly developing wing of experimental economics shows precisely the same tendency. Both fields often neglect the problem of generalization and thereby invite the worry that their results and theories will not travel very far beyond their specialized settings and subjects.

While successful experimentation in political science must acknowledge the threats to generalization that center on settings, populations, and treatments, there is much by a way of remedy that can be done. The most effective response to the problem of experimental generalization is to carry out *selective* replications. We emphasize selective because we have no interest, and the field has nothing like the required resources, to follow a program of *comprehensive* replication. By no means are we suggesting that, to be assured of the generalizability of a particular experimental result, we must have corroborating evidence from representative samples of populations, representative samples of settings, and representative samples of treatments. Such a recommendation is foolish, and not least because, were it to be enforced, it would quickly persuade those with creative ideas and experimental inclinations to take up other lines of work.

Our advice, instead, is to pursue carefully chosen and selective replications. The point is to vary settings, populations, and treatments in ways that represent revealing and usually difficult tests of generalization. In this way, we *probe* the generalizability of experimental results. For reasons of control and convenience, most experiments will no doubt continue to be undertaken in artificial settings with college student subjects confronting treatments that have no exact counterparts outside the laboratory. But these must be complemented by occasional experimental ventures that place a higher premium on matters of external validity.

10. Campbell has softened his position in recent years; compare Campbell and Stanley 1966 and Campbell 1969a with Cook and Campbell 1979.

Happily, there appears to be a fair number of such experiments already underway in political science. For example, a number of experiments in political science have been conducted in the field, in natural settings, and unobtrusively, so that the possibility of subjects reacting to the artificial aspects of the research setting simply never arises (Campbell 1969a; Ellsworth 1977). Thus, Cover and Blumberg (1982 and Part 4) investigate the celebrated incumbency advantage enjoyed by members of the U.S. House of Representatives by withholding the flow of congressional mail from some constituents and not from others. From the constituent subject's point of view, the experiment is utterly innocuous. Similarly, it seems most unlikely that Chicagoans who happened to receive a post card in the mail emphasizing the citizen's sacred obligation to vote suspected that they were being experimented upon by Dr. Gosnell (or anyone else). Or consider Levine and Plott's (1977 and Part 5) demonstration of just how effective an agenda setter can be at influencing committee decisions. Such power had been suggested by several theoretical models from social choice and game theory and had some casual support from descriptive empirical observation. Yet the theoretical limits of this power had never been measured in a careful, systematic way, and so there was good reason for skepticism. Could highly simplified and abstract formal theories really predict outcomes in complicated committees? Levine and Plott's *field* experiment of agenda influence in the context of a flying club deciding on a fleet of airplanes provide strong evidence in favor of theories that claim that agenda manipulations are highly predictable from only limited information about the underlying process and environment.[11] Finally, consider Fiorina and Plott's result (1978 and Part 5) that the process of committee decision making takes a very different path when real, material incentives are at stake. This is a prime example of a selective replication with a message, and the message is that experimental studies on group decision making that ignore material incentives do so at their peril. Twenty years earlier, Sidney Siegel showed how to convert this thorny criticism about experiments into a startling and brilliant contribution to theories of motivation and decision making (Siegel 1961). The point here is again that experimental methodology has the flexibility to respond effectively and insightfully to many naive criticisms. Usually the response is to enrich the experimental design—a luxury that other empirical methodologies do not have available to them.

Although college students dominate experiments in political science, as they do experiments throughout the social sciences, it is not difficult to find

11. For another example, similar in spirit, see Romer and Rosenthal's (1979) research on the strategic manipulation of school budget referenda.

experiments motivated by political concerns that reach beyond the borders of the campus. Two obvious and exemplary illustrations are the experiments conducted by Cover and Blumberg and by Gosnell (noted previously). Both these experiments not only went beyond the campus but settled on exactly the populations of real interest. Consider also the torrent of experimental work devoted to understanding the nature of public opinion, accomplished by embedding manipulations of question wording, question order, or question frame, into sample surveys (e.g., Bradburn 1982; Schuman and Presser 1980 and 1981; Sullivan, Piereson, and Marcus 1978 and Part 1; Tourangeau and Rasinski 1988). Currently, this is a very lively area of research, supported by both the National Academy of Sciences and the Social Science Research Council, and contributed to by cognitive psychologists who are attracted to surveys as an alternative research site to the university laboratory, and by survey researchers who want assistance in understanding how people answer questions.[12]

Identifying the population of real interest and persuading members of that population to participate—as in Weiss's (1982) experimental study of complexity and decision making—does not fully resolve the problem of subject generalizability, of course; some form of selection bias may remain. As Webb and his colleagues put it, a touch dramatically, the "curious, the exhibitionistic, and the succorant" are likely to show up in disproportionate numbers in any sample of volunteers (Webb et al. 1966, 25). Nevertheless, it is quite easy to decide that we would rather to see grown-up policy analysts run through Weiss's experiment than the garden variety college sophomore.

Finally, a standard and useful response to the problem posed by the generalizability of treatments is to create experimental manipulations that fall within the typical range. In the Iyengar and Kinder television experiments, for example, manipulations were designed to recreate typical broadcasts. It would have been easy, and not particularly interesting, to demonstrate television effects by burying participants under an avalanche of stories about a particular problem. Such a demonstration would be revealing about the potential power of television under extraordinary circumstances. Iyengar and Kinder's interest was, rather, in the realized power of television under ordinary circumstances, and so experimental treatments were designed to fall within the typical range of television news.

This kind of ecological representation is not always appropriate, however. A nice exception is provided by McKelvey and Ordeshook's election experiments. Such experiments place severe restrictions on what voters know about the choices they confront, restrictions that surely surpass naturally occurring restrictions. If, under such circumstances, which obtain nowhere outside the artificial world that McKelvey and Ordeshook create, voters nevertheless behave in ways that live up to the expectations set by full information models, then, as they say, "we feel confident in concluding that lack of information, by itself, does not necessarily preclude democratic process from being attracted to full information outcomes" (1990, 312). Extreme experimental manipulations can provide provocative insights on real political processes.

Taken together, these varied examples suggest that the problem of generalization, while real enough, is no reason to give up on experiments. The risks run in generalizing from experimental results may never be eliminated entirely, but they can be sharply reduced: by diminishing, or bypassing altogether, the artificiality of the experimental setting; by extending experimental tests to diverse or difficult samples; by creating treatments that are ecologically representative; and more. Through ingenuity, opportunism, and sheer effort, the "scandal of induction" becomes just another research challenge, no different in kind from the altogether familiar everyday problems of design, measurement, and analysis.

5. A Look Ahead

Acknowledging the limitations of experiments, as we have just done, does no serious damage to our central contention that experimentation deserves to be taken more seriously in political science. All methods have their shortcomings, including the archival and sample survey techniques that have been political science's mainstays over the last half-century. And in the case of experimentation, the shortcomings are vastly outweighed by the advantages that experimentation provides. Testing causal propositions, decomposing complex forces, accelerating interdisciplinary conversations, turning up replicable "stubborn facts," moving smoothly across different levels of aggregation: the advantages of experimentation are—or should be—irresistible. Political science needs all the help it can get; it needs a full arsenal of empirical methods, experimentation chief among them, as it makes its way, haltingly, "From symbols and shadows to the truth."[13]

Exhortations, such as the one we have just delivered, are often just so many words, intended more for the pleasure of the writer than the reader. We hope that this volume does more than make us feel virtuous. To be persuasive about the advantages of experimentation probably requires more than we have done so far: it requires showing off the theoretical merits of experimentation

12. This sounds good, and it is. Notice, however, that while inserting experiments into probability sample surveys is a powerful solution to the standard complaint about relying too much on college sophomores, it has nothing to say to the parallel complaints about settings (after all, surveys take place in a specialized setting, typically in the respondent's living room) or treatments (since such experiments are limited typically to the manipulation of semantic material).

13. The passage comes from Cardinal Newsman's epitaph, quoted in Webb et al. 1966, 185.

with real, practical examples. With this in mind, the rest of the book proceeds to cook up and serve an experimental banquet: a feast of delicious experiments applied to the study of politics. These experiments are not perfect, but they are exemplary in displaying experimental advantages, and they have made a real difference in how we think about important aspects of politics.

The 15 articles we have chosen to reprint here cover a wide range of substantive topics, organized around five themes. First is a set of experiments that focus on public opinion, examining how individual citizens attempt to make sense of what Lippmann (1922) called "the mystery out there." Part 2 is devoted to experiments on political choice, inspired for the most part by recent discoveries of systematic departures from what is usually thought of as fully rational decision making. Part 3 comprises a set of representative experiments from the interdisciplinary research effort on collective action, where individual incentives collide with the common good. The fourth part takes up, again from an experimental perspective, the problem of information in democracy: how and how well do citizens learn about politics and what difference does it make for governance? Finally, Part 5 presents a series of experiments that investigate how variations in institutional rules and procedures affect governmental outcomes.

The topical diversity of what is to come is remarkable, which can be counted as both an experimental virtue and an aesthetic liability. To make certain that the former dominates the latter, we introduce each of the five sections with a brief essay that places the experimental work in context and provides suggestions for further exploration. We hope that the volume as a whole will be experienced as a full and satisfying meal: that the strengths of experimentation set out in this essay will come fully to life in the detailed revelations that follow. Actually, we would prefer if the experience were not completely satisfying: our real ambition is to stir up an appetite for further experimentation. Think of the book then as an hors d'oeuvre that will leave you hungry for more.

6. Guide to Additional Readings

Our essay necessarily leaves aside some important topics. In partial compensation, we offer here a few suggestions for further reading for those who wish to go deeper into the experimental method. First, on the logic of experimentation, the place to begin is Campbell and Stanley's *Experimental and Quasi-Experimental Designs for Research* (1966). Nobody does it better, with the possible exception of Cook and Campbell's *Quasi-Experimentation* (1979). On the pragmatics of doing experiments—a primer for action—the best single source is *Methods of Research in Social Psychology* by Carlsmith, Ellsworth, and Aronson (1976). On the statistical analysis of experimental data, consult Kenny (1985) and Winer (1971). The basic principles underlying the use of experiments to evaluate formal theories of political processes are well articulated in Plott (1979). One of those principles is that preferences can be experimentally induced by monetary rewards, as discussed in Smith (1976).

Treat *Methods of Research in Social Psychology* as a companion volume to *Experimental and Quasi-Experimental Designs for Research:* read them together. The Campbell and Stanley book is enormously useful, but not exactly inspirational. After pages and pages of threats to valid inference, described in menacing vocabulary ("instrument decay," "subject mortality"), the intrepid reader is likely to emerge with a heavy heart. Consider this passage, which comes as close as Campbell and Stanley get to a sermon on experiment's behalf.

> If, as seems likely, the ecology of our science is one in which there are available many more wrong responses than correct ones, we may anticipate that most experiments will be disappointing. We must somewhat inoculate young experimenters against this effect, and in general must justify experimentation on more pessimistic grounds—not as a panacea, but rather as the only available route to cumulative progress. We must instill in our students the expectation of tedium and disappointment and the duty of thorough persistence, by now so well achieved in the biological and physical sciences. We must expand our students' vow of poverty to include not only the willingness to accept poverty of finances, but also a poverty of experimental results. (1966, 3)

Not exactly a call to arms! And not an accurate account of the life of an experimenter, at least as we have lived it. In writing to protect against disappointment, Campbell and Stanley went too far: we worry that their dark sentiments may have prevented disillusionment primarily by discouraging experimental work from getting off the ground in the first place.

Carlsmith, Ellsworth, and Aronson offer quite a different and upbeat introduction to experimentation. *Methods of Research in Social Psychology* is an attempt, largely successful, to demystify experimental practice. They provide a thoughtful, how-to guide that will prove indispensable to anyone contemplating undertaking an experiment, and they convey the sense, which we share, that the planning, execution, analysis, and interpretation of experiments is often great fun.

REFERENCES

Abelson, Robert P., and Ariel Levi. 1985. Decision Making and Decision Theory. In Gardner Lindzey and Elliot Aronson, eds., *Handbook of Social Psychology*. New York: Random House.

Alesina, Alberto. 1987. Macroeconomic Policy in a Two-Party System as a Repeated Game. *Quarterly Journal of Economics* 102:651–78.

Aronson, Elliot, and Judson Mills. 1959. The Effect of Severity of Initiation on Liking for a Group. *Journal of Abnormal and Social Psychology* 59:177–81.

Austen-Smith, David, and Jeffrey Banks. 1989. Electoral Accountability and Incumbency. In Peter Ordeshook, ed., *Models of Strategic Choice in Politics*. Ann Arbor: University of Michigan Press.

Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic Books.

Baron, David, and John Ferejohn. 1989. The Power to Propose. In Peter Ordeshook, ed., *Models of Strategic Choice in Politics*. Ann Arbor: University of Michigan Press.

Boring, Edwin G. 1954. The Nature and History of Experimental Control. *American Journal of Psychology* 67:573–89.

Boylan, Richard, John Ledyard, Arthur Lupia, Richard D. McKelvey, and Peter Ordeshook. 1991. Political Competition in a Model of Economic Growth: An Experimental Study. In Thomas R. Palfrey, ed., *Laboratory Research in Political Economy*. Ann Arbor: University of Michigan Press.

Bradburn, Norman. 1982. Question-Wording Effects in Surveys. In Robert Hogarth, ed., *Question Framing and Response Consistency*. San Francisco: Jossey-Bass.

Brody, Richard A., and Charles N. Brownstein. 1975. Experimentation and Simulation. In F. I. Greenstein and N. Polsby, eds., *Handbook of Political Science* 7:211–63. Reading, Mass.: Addison-Wesley.

Burnham, Walter D. 1970. *Critical Elections and the Mainsprings of American Politics*. New York: Norton.

Campbell, Angus, Philip E. Converse, Warren E. Miller, and Donald E. Stokes. 1960. *The American Voter*. New York: Wiley.

Campbell, Donald T. 1969a. Prospective: Artifact and Control. In Robert Rosenthal and Robert Rosnow, eds., *Artifact in Behavioral Research*. New York: Academic Press.

Campbell, Donald T. 1969b. Reforms as Experiments. *American Psychologist* 24:409–29.

Campbell, Donald T., and H. L. Ross. 1970. The Connecticut Crackdown on Speeding: Time-Series Data in Quasi-Experimental Analysis. In Edward R. Tufte, ed., *The Quantitative Analysis of Social Problems*. Reading, Mass.: Addison-Wesley.

Campbell, Donald T., and Julian C. Stanley. 1966. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.

Carlsmith, J. Merrill, Phoebe C. Ellsworth, and Elliot Aronson. 1976. *Methods of Research in Social Psychology*. Reading, Mass.: Addison-Wesley.

Chamberlin, John R. 1978. The Logic of Collective Action: Some Experimental Results. *Behavioral Science* 23:441–45.

Chew, S. H. 1983. A Generalization of the Quasilinear Mean with Applications to the Measurement of Income Inequality and Decision Theory Resolving the Allais Paradox. *Econometrica* 51:1065–92.

Clubb, Jerry M., William H. Flanigan, and Nancy H. Zingale. 1980. *Partisan Realignment*. Beverly Hills: Sage.

Cohen, Bernard. 1963. *The Press and Foreign Policy*. Princeton: Princeton University Press.

Collier, Kenneth E., Richard D. McKelvey, Peter C. Ordeshook, and Kenneth C. Williams. 1987. Retrospective Voting: An Experimental Study. *Public Choice* 53:101–30.

Converse, Philip E. 1964. The Nature of Belief Systems in Mass Publics. In David E. Apter, ed., *Ideology and Discontent*. New York: Free Press.

Converse, Philip E. 1975. Public Opinion and Voting Behavior. In Fred Greenstein and Nelson Polsby, eds., *Handbook of Political Science*. Vol. 4. Reading, Mass.: Addison-Wesley.

Converse, Philip E. 1990. Popular Representation and the Distribution of Information. In John Ferejohn and James Kuklinski, eds., *Information and Democratic Processes*. Urbana-Champaign: University of Illinois Press.

Cook, Thomas D., and Donald T. Campbell. 1979. *Quasi-Experimentation*. Chicago: Rand McNally.

Cover, Albert D., and Bruce S. Brumberg. 1982. Baby Books and Ballots: The Impact of Congressional Mail on Constituent Opinion. *American Political Science Review* 76:347–59.

Cukierman, Alex. 1985. Asymmetric Information and the Electoral Momentum of Public Opinion Polls. Presented at the Weingart Conference on Formal Models in Voting, California Institute of Technology.

Darwin, Charles. 1875. *Insectivorous Plants*. London: J. Murray.

Darwin, Charles, and Francis Darwin. 1890. *The Power of Movement in Plants*. New York: Appleton.

Dawes, Robyn M. 1980. Social Dilemmas. *Annual Review of Psychology* 31:169–93.

Dawes, Robyn M., John M. Orbell, Randy T. Simmons, and Alphons J. C. van de Kragt. 1986. Organizing Groups for Collective Action. *American Political Science Review* 80:1171–85.

Dekel, E. 1986. An Axiomatic Characterization of Preferences under Uncertainty: Weakening the Independence Axiom. *Journal of Economic Theory* 40:304–18.

Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper and Row.

Eavey, Cheryl L., and Gary J. Miller. 1984. Bureaucratic Agenda Control: Imposition or Bargaining? *American Political Science Review* 78:719–33.

Ellsworth, Phoebe C. 1977. From Abstract Ideas to Concrete Instances: Some Guidelines for Choosing Natural Research Settings. *American Psychologist* 32:604–15.

Fenno, Richard F., Jr. 1978. *Home Style: House Members in Their Districts*. Boston: Little, Brown.

Ferejohn, John, Robert Forsythe, Roger Noll, and Thomas R. Palfrey. 1982. An Experimental Examination of Auction Mechanisms for Discrete Public Goods. In Vernon L. Smith, ed., *Research in Experimental Economics* 2:175–99. Greenwich, Conn.: JAI Press.

Fiorina, Morris P. 1981. *Retrospective Voting in American National Elections*. New Haven: Yale University Press.

Fiorina, Morris P., and Charles R. Plott. 1978. Committee Decisions Under Majority Rule. *American Political Science Review* 72:575–98.

Fisher, Ronald. 1935. *Design of Experiments*. New York: Hafner Publishing.

Gerard, Harold B., and G. C. Mathewson. 1966. The Effects of Severity of Initiation on Liking for a Group: A Replication. *Journal of Experimental Social Psychology* 2:278–87.

Gosnell, Harold F. 1927. *Getting Out the Vote: An Experiment in the Stimulation of Voting.* Chicago: University of Chicago Press.

Grether, David M. 1980. Bayes' Rule as a Descriptive Model: The Representativeness Heuristic. *Quarterly Journal of Economics* 95:537–57.

Grether, David M., and Charles Plott. 1979. Economic Theory of Choice and the Preference Reversal Phenomenon. *American Economic Review* 69:623–38.

Grunberg, E., and F. Modigliani. 1954. The Predictability of Social Events. *Journal of Political Economy* 62:465–78.

Herstein, John A. 1981. Keeping the Voter's Limits in Mind: A Cognitive Process Analysis of Decision Making in Voting. *Journal of Personality and Social Psychology* 40:843–61.

Hovland, Carl I. 1959. Reconciling Conflicting Results Derived from Experimental and Survey Studies of Attitude Change. *American Psychologist* 14:8–17.

Isaac, R. Mark, James M. Walker, and Susan H. Thomas. 1984. Divergent Evidence on Free Riding: An Experimental Examination of Possible Explanations. *Public Choice* 43:113–49.

Iyengar, Shanto, and Donald R. Kinder. 1987. *News that Matters.* Chicago: University of Chicago Press.

Iyengar, Shanto, Mark D. Peters, and Donald R. Kinder. 1982. Experimental Demonstrations of the "Not-So-Minimal" Consequences of Television News Programs. *American Political Science Review* 76:848–58.

Jervis, Robert. 1976. *Perceptions and Misperception in International Relations.* Princeton: Princeton University Press.

Kagel, John, Ronald Harstad, and Dan Levin. 1987. Information Impact and Allocation Rules in Auctions with Affiliated Private Values: A Laboratory Study. *Econometrica* 55:1275–1304.

Kahneman, Daniel, Paul Slovic, and Amos Tversky. 1982. *Judgment under Uncertainty: Heuristics and Biases.* Cambridge: Cambridge University Press.

Kahneman, Daniel, and Amos Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47:263–91.

Kaplan, Abraham. 1964. *The Conduct of Inquiry.* San Francisco: Chandler Publishing.

Kenny, David A. 1985. Quantitative Methods for Social Psychology. In Gardner Lindzey and Elliot Aronson, eds., *Handbook of Social Psychology.* New York: Random House.

Key, V. O. 1949. *Southern Politics.* New York: Knopf.

Key, V. O., Jr. 1955. A Theory of Critical Elections. *Journal of Politics* 17:3–18.

Key, V. O., Jr. 1966. *The Responsible Electorate.* Cambridge, Mass.: Harvard University Press.

Kinder, Donald R., and Lynn M. Sanders. 1990. Mimicking Political Debate with Survey Questions: The Case of White Opinion on Affirmative Action for Blacks. *Social Cognition* 8:73–103.

Kinder, Donald R., and David O. Sears. 1985. Public Opinion and Political Action. In Gardner Lindzey and Elliot Aronson, eds., *Handbook of Social Psychology.* New York: Random House.

Knez, Marc, and Vernon L. Smith. 1987. Hypothetical Valuations and Preference Reversals in the Context of Asset Trading. In Alvin Roth, ed., *Laboratory Experimentation in Economics.* Cambridge: Cambridge University Press.

Kramer, Gerald H. 1971. Short-Term Fluctuations in U.S. Voting Behavior, 1896–1964. *American Political Science Review* 65:131–43.

Kydland, Finn, and Edward Prescott. 1977. Rules Rather than Discretion: The Inconsistency of Optimal Plans. *Journal of Political Economy* 85:473–91.

Lane, Robert E. 1962. *Political Ideology.* New York: Free Press.

Lazarsfeld, Paul F., and Robert K. Merton. 1948. Mass Communication, Popular Taste, and Organized Social Action. In L. Bryson, ed., *The Communication of Ideas.* New York: Harper.

Ledyard, John O. 1989. Information Aggregation in Two-Candidate Elections. In Peter Ordeshook, ed., *Models of Strategic Choice in Politics.* Ann Arbor: University of Michigan Press.

Levine, Michael E., and Charles R. Plott. 1977. Agenda Influence and Its Implications. *Virginia Law Review* 63:561–604.

Lodge, Milton, and Ruth Hamill. 1986. A Partisan Schema for Political Information Processing. *American Political Science Review* 80:505–19.

Lippmann, Walter. 1922. *Public Opinion.* New York: Macmillan.

Lippmann, Walter. 1925. *The Phantom Public.* New York: Harcourt, Brace.

Machina, Mark. 1982. "Expected Utility" Analysis without the Independence Axiom. *Econometrica* 50:277–323.

Machina, Mark. 1983. The Economic Theory of Individual Behavior toward Risk: Theory, Evidence and New Directions. Working Paper. Institute for Mathematical Studies in the Social Sciences. Stanford University.

McKelvey, Richard D., and Peter C. Ordeshook. 1984. Rational Expectations and Elections: Some Experimental Results Based on a Multidimensional Model. *Public Choice* 44:61–102.

McKelvey, Richard D., and Peter C. Ordeshook. 1985. Elections with Limited Information: A Fulfilled Expectations Model Using Contemporaneous Poll and Endorsement Data as Information Sources. *Journal of Economic Theory* 36:55–85.

McKelvey, Richard D., and Peter C. Ordeshook. 1990. A Decade of Experimental Research on Spatial Models of Elections and Committees. In James Enelow and Melvin Hinich, eds., *Advances in the Spatial Theory of Voting.* Cambridge: Cambridge University Press.

McKelvey, Richard D., and Peter C. Ordeshook. 1990. Information and Elections: Retrospective Voting and Rational Expectations. In John Ferejohn and James Kuklinski, eds., *Information and Democratic Processes.* Urbana-Champaign: University of Illinois Press.

Marwell, Gerald, and R. Ames. 1980. Experiments on the Provision of Public Goods II: Provision Points, Stakes, Experience and the Free Rider Problem. *American Journal of Sociology* 85:926–37.

Marwell, Gerald, and R. Ames. 1981. Economists Free Ride, Does Anyone Else? Experiments on the Provision of Public Goods IV. *Journal of Public Economics* 15:295–310.

Milgram, Stanley. 1974. *Obedience to Authority.* New York: Harper and Row.

Muth, John F. 1961. Rational Expectations and the Theory of Price Movements. *Econometrica* 29:315–35.

Nie, Norman H., Sidney Verba, and John R. Petrocik. 1979. *The Changing American Voter.* Cambridge, Mass.: Harvard University Press.

Nordhaus, William. 1975. The Political Business Cycle. *Review of Economic Studies* 42:169–90.

Olson, Mancur. 1965. *The Logic of Collective Action.* Cambridge, Mass.: Harvard University Press.

Ordeshook, Peter, and Thomas R. Palfrey. 1988. Agendas, Strategic Voting, and Signaling with Incomplete Information. *American Journal of Political Science* 32:441–66.

Page, Benjamin I., and Robert Y. Shapiro. 1992. The Rational Public and Democracy. Chicago: University of Chicago Press.

Palfrey, Thomas R., and Howard Rosenthal. 1984. Participation and Provision of Discrete Public Goods: A Strategic Analysis. *Journal of Public Economics* 24:171–93.

Palfrey, Thomas R., and Howard Rosenthal. 1988. Private Incentives in Social Dilemmas: The Effects of Incomplete Information and Altruism. *Journal of Public Economics* 35:309–32.

Plott, Charles R. 1979. The Application of Laboratory Experimental Methods to Public Choice. In C. Russell, ed., *Collective Decision Making.* Baltimore: Johns Hopkins University Press.

Quattrone, George A., and Amos Tversky. 1988. Contrasting Rational and Psychological Analyses of Political Choice. *American Political Science Review* 82:719–36.

Rice, S. A. 1929. Contagious Bias in the Interview: A Methodological Note. *American Journal of Sociology* 35:420–23.

Roberts, J. 1976. The Incentives for the Correct Revelation of Preferences and the Number of Consumers. *Journal of Public Economics* 6:359–74.

Rogoff, Kenneth, and Anne Sibert. 1988. Elections and Macroeconomic Policy Cycles. *Review of Economic Studies* 55:1–16.

Romer, T., and H. Rosenthal. 1979. Bureaucrats Versus Voters: On the Political Economy of Resource Allocation by Direct Democracy. *Quarterly Journal of Economics* 94:563–87.

Samuelson, Paul. 1954. The Price Theory of Public Expenditure. *Review of Economics and Statistics* 36:387–89.

Samuelson, William F., and Max H. Bazerman. 1985. The Winner's Curse in Bilateral Negotiations. In Vernon Smith, ed., *Research in Experimental Economics* 3:104–37. Greenwich, Conn.: JAI Press.

Schuman, Howard, and Stanley Presser. 1980. Public Opinion and Public Ignorance: The Fine Line between Attitudes and Nonattitudes. *American Journal of Sociology* 85:1214–25.

Schuman, Howard, and Stanley Presser. 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Wording, Form and Context.* New York: Academic Press.

Sears, David O. 1986. College Sophomores in the Laboratory: Influence of a Narrow Data Base on Social Psychology's View of Human Nature. *Journal of Personality and Social Psychology* 51:515–30.

Siegel, Sidney. 1961. Decision Making and Learning under Varying Conditions of Reinforcement. *Annals of the New York Academy of Science,* 766–83.

Simon, Herbert A. 1954. Bandwagon and Underdog Effects of Election Predictions. *Public Opinion Quarterly* 18:245–54.

Smith, Vernon L. 1976. Experimental Economics: Induced Value Theory. *American Economic Review* 66:274–79.

Smith, Vernon L. 1982. Microeconomic Systems as an Experimental Science. *American Economic Review* 72:923–55.

Sullivan, John L., James E. Piereson, and George E. Marcus. 1978. Ideological Constraint in the Mass Public: A Methodological Critique and Some New Findings. *American Journal of Political Science* 22:233–49.

Sundquist, James L. 1973. *Dynamics of the Party System.* Washington, D.C.: Brookings.

Thaler, R. 1980. Toward a Positive Theory of Consumer Choice. *Journal of Economic Behavior and Organization* 1:39–60.

Thompson, Dennis F. 1970. *The Democratic Citizen.* Cambridge: Cambridge University Press.

Tourangeau, Roger, and Kenneth A. Rasinski. 1988. Cognitive Processes Underlying Context Effects in Attitude Measurement. *Psychological Bulletin* 103:299–314.

Tversky, Amos, and Daniel Kahneman. 1981. The Framing of Decisions and the Psychology of Choice. *Science* 211:453–58.

Walker, Jack L. 1983. The Origins and Maintenance of Interest Groups in America. *American Political Science Review* 77:392–406.

Webb, Eugene J., Donald T. Campbell, Richard D. Schwartz, and Lee Sechrest. 1966. *Unobtrusive Measures: Nonreactive Research in the Social Sciences.* Chicago: Rand McNally.

Weiss, Janet. 1982. Coping with Complexity: An Experimental Study of Public Policy Decision Making. *Journal of Policy Analysis and Management* 2:66–87.

Winer, B. J. 1971. *Statistical Principles in Experimental Design.* New York: McGraw-Hill.

Zaller, John. 1985. Toward a Theory of the Survey Response. Department of Political Science, University of California, Los Angeles. Photocopy.