Experimental Methodology in Political Science

Author(s): Rose McDermott

Source: *Political Analysis*, Autumn 2002, Vol. 10, No. 4, Special Issue on Experimental Methods in Political Science (Autumn 2002), pp. 325-342

Published by: Cambridge University Press on behalf of the Society for Political Methodology

Stable URL: https://www.jstor.org/stable/25791696

Experimental Methodology in Political Science

Rose McDermott
Department of Government, McGraw Hall,
Cornell University, Ithaca, NY 14850
e-mail: rmm21@cornell.edu

Experiments offer a useful methodological tool to examine issues of importance to political scientists. The historical and cultural differences between experiments in behavioral economics and social psychology are discussed. Issues of central concern to experimentalists are covered, including impact versus control, mundane versus experimental realism, internal versus external validity, deception, and laboratory versus field experiments. Advantages and disadvantages of experimentation are summarized.

1 Introduction

A few years ago, several physicians associated with the National Bowel and Breast Cancer multisite trials lost their jobs and reputations as well as millions of dollars in research funding because of their failure to adhere properly to the experimental protocols of a study on lumpectomy versus mastectomy in breast cancer patients. In particular, one of the participating physicians in Toronto was admitting patients whose tumors were larger than the accepted protocol for study and treatment. When asked about this behavior, the physician replied that he was simply trying to help his patients by giving them access to state-of-the-art treatment. Yet his violation of protocol, along with that of others, forced the complete reanalysis of data that had previously indicated that lumpectomy with radiation was as effective as mastectomy in treating certain breast cancers. Although the overall recommendations did not shift as a result of this reexamination, thousands of women were unnecessarily terrified and many more people lost faith in the reliability and honesty of medical research. Why did this happen? Because a few smart and medically well-trained doctors failed to understand the important methodological requirements of proper experimental procedure.

This article addresses the proper purpose and use of experimental methods within political science. It begins with a discussion of the differences among experiments in behavioral economics, social psychology, and political science. It then addresses some of the main methodological issues that arise in connection with the use of experimental procedure. A brief discussion of the advantages and disadvantages of using experiments follows.

Author's note: The author thanks Arthur Lupia, an anonymous reviewer, and James Alt and the participants at the Experimental Conferences at the Center for Basic Research in the Social Sciences at Harvard University for very helpful suggestions and comments throughout. I also thank the psychologists who helped to train me in experimental methods and design: Philip Zimbardo, Lee Ross, and the late Amos Tversky.

I do not argue that experimentation is the only, or even the best, methodological form of investigation, but rather that it is one valuable tool among many possible methods, and a strategy that should not be neglected out of hand. Experiments may not be useful for all research questions or problems, but can be especially helpful to investigators who have had difficulty achieving consensus using other methods, seek a stronger basis for causal argumentation, or are interested in developing and testing theoretical models in a direct, empirically grounded manner. Also, because of the unique benefits offered by experimentation, failure to use this strategy may impose unwarranted limitations on the state of knowledge and the speed of its development in our field.

1.1 *History*

Informal experimentation goes back at least as far as the court members surrounding French King Louis XIV, who began to develop probability theory in an attempt to improve their winnings in the gambling games that consumed much of their free time. Probably the earliest published example of an experimental study reported Bernoulli's (1738) investigations of the St. Petersburg paradox, which demonstrated certain limitations of expected utility as a universal normative model.

Systematic experimentation has produced a long and impressive series of findings in most disciplines in the hard sciences, including biology, chemistry, and physics. This is true in the social and behavioral sciences as well. Experimentation in psychology goes back at least to the end of the 19th century. Since the mid-1980s, in particular, work in behavioral economics has transformed the broader field of economics through the systematic use of experimentation. Experiments in social psychology and behavioral economics often focus on issues of concern to political scientists. However, the history and purposes of each of these disciplines differ enough that they have produced divergent styles and goals of experimentation that often lead to confusion when experimentally naive observers attempt to interpret or integrate them. It is thus worth reviewing briefly the different traditions of each field before proceeding to a discussion of the relevant methodological issues that challenge any experimenter.

1.1.1 Behavioral Economics[1]

One of the most diagnostic features of economic experiments is the monetary incentives that are used to induce and validate subjects' behavior. In some ways, this makes sense given the economic topics under investigation. Yet in another way, there is no empirical reason to assume a priori that money ensures that subjects will behave more closely in line with experimental instructions simply because there is a fiscal reward for behaving rationally in the experiment. Individuals can have other, particularly powerful social, goals as well. Experimental biases that plague noneconomic experiments, including demand characteristics, social desirability biases, and so on, might still operate even in the face of monetary incentives.

The reason for the widespread use of monetary incentives in economic experiments is at least partly historical as much as theoretical. An early experiment by Thurstone (1931), designed to develop individual indifference curves, was criticized by W. Allen Wallis and Milton Friedman (1942) for the hypothetical nature of the choices presented to the subjects.

[1]This section draws on the first part of the introduction chapter by Alvin Roth of John Hagel and Alvin Roth (Eds.), *The Handbook of Experimental Economics,* 1995.

Wallis and Friedman argued that in order to learn anything useful from an experiment, subjects needed to react to "actual stimuli" and not merely to their conjectures of how they might behave if they were confronted by a real-life situation such as the one described experimentally. This criticism gained widespread credence among future experimentalists, and monetary payoffs became the *sine qua non* of economic experimentation. Monetary rewards allowed economists to assume that subjects were equalized by their striving to make the most money possible in all experimental situations. In addition, transparency became the name of the game as subjects were expected to be able to see and understand how the monetary rewards worked in order for the monetary incentives to have an effect. Experimenters wanted subjects to realize that their reward was contingent on their ostensibly rational behavior.

Early economic theorists such as Adam Smith and John Maynard Keynes understood that humans were psychologically complex and incorporated this understanding into their models. However, by the 1920s economists were striving to mold themselves into a "real" science, like physics. This required complex mathematical models dependent on assumptions of people as rational, self-interested actors who respond to price changes but not to emotions in making their economic decisions (Uchitelle 2001). By the 1940s, this notion of "rational man" formed the orthodoxy of economics. This model was advanced by the increasing use of mathematics in economics. Of course, mathematical models are much easier to formulate when people are assumed to be rational (*Economist* 2000).

However, following the publication of von Neumann and Morganstern's (1944) seminal work *Theory of Games and Economic Behavior*, experimental economics began to focus more on individual choice, expected utility, and game theory. These topics seemed ideal for experimental investigation because they allowed scholars to explore individual behavior under controlled conditions that were unaffected by extraneous contaminants. Even Thomas Schelling wrote in 1958 that "the principles relevant to *successful* play, the *strategic* principles, the propositions of *normative* theory, cannot be derived by purely analytic means from a priori considerations" (p. 257). Among the early enthusiasts for this method were Melvin Dresher and Merrill Flood, who began a series of experiments at the Rand Corporation in 1950 that subsequently came to be known as the Prisoner's Dilemma paradigm.

By the early 1950s, economists had begun to concentrate on using experiments to test theories, especially game theories, in controlled environments. Economists sought to make clear predictions and test them in a variety of stripped-down settings. When the predictions did not work out, economists tested alternative hypotheses in subsequent work. Anonymity of subjects came to be important as investigators sought to remove extraneous factors that might contaminate the experimental findings. Building on the insights gained from successes, failures, and anomalies in earlier experimental work, economists began an ongoing interaction between theoretical advances and experimental empirical investigation designed to test and develop formal models of human behavior.

Experimental work continued to expand throughout the 1960s. The first large funding sources for experimental economics began in the 1970s, when the National Science Foundation (NSF) began to provide systematic and ongoing support to a few laboratories throughout the country. Importantly, Eric Wanner at the Russell Sage Foundation began to fund Richard Thaler's work on anomalies in human economic behavior, work that threatened the dominant economic model of man as a rational actor (Lowenstein 2001). It was at this time that experimental economics began to diverge increasingly from experimental psychology, although there continues to be a great deal of overlap, especially in work on individual choice behavior.

Experimental economics really arrived when, in 1988, Maurice Allais, who had discovered an important paradox that revealed a way in which individuals systematically deviate from rational behavior in choice, won the Nobel Prize in economics. By this time, several researchers in behavioral economics, most notably Richard Thaler and George Akerlof, began to establish systematic programs of research in behavioral economics that led to the replication and cumulation of knowledge around several specific substantive areas. Interestingly, both men were heavily influenced by psychologists Amos Tversky and Daniel Kahneman, who themselves pioneered work on anomalies in human judgment and decision making.

Two other major breakthroughs in experimental economics occurred around this time as well. First, the stock market crashed 23% in one day on October 19, 1987, destroying the beliefs of many economists in the triumph of rational actors operating in efficient markets (Lowenstein 2001). Interest strengthened and developed in the work of Thaler and others whose experimental work demonstrated ways in which individual behavior systematically deviated from rationalist expectations. A second breakthrough in the history of behavioral economics occurred when Harvard hired David Laibson in 1994, signaling the first time that a major institution had deliberately hired a behavioral economist (Uchitelle 2001).

Currently, experiments in behavioral economics deemed to be successful share certain core characteristics with precedents that should now be obvious: clear instructions; absolutely no deception; stylized stripped down settings; anonymity; cash incentives; an experimental test of a formal model, which can include a comparison with rational models; and, ideally, subsequent validation of the findings with supportive field data (Laibson 2000).

The substantive topics that have most preoccupied behavioral economics research include public goods, coordination problems, bargaining, experimental market behavior, auctions, and individual decision making (Kagel and Roth 1995). Clearly, these issues overlap with issues that preoccupy political scientists. In particular, many of these topics relate closely to topics investigated by political economists experimentally: strategic behavior, incomplete and/or asymmetric information, and the dynamics of change (Palfrey 1991). These areas of overlap might easily be exploited experimentally for the mutual benefit of both disciplines.

In 1959, economists Siegel and Fouraker (1960) won the Monograph Prize of the American Academy of Arts and Sciences with an experiment investigating levels of aspiration. Their article, as noted by Kagel and Roth (1995), makes an argument for the virtues of the experimental method that continues to hold true today:

> Our data have been observations made specifically to meet the purposes of this research. We have not turned to preexisting data . . . .
>
> It is when a science starts to go out and ask exactly what it wants to know, rather than relying on information collected for other purposes, that it begins to obtain control of its own growth. (pp. 72–73)

1.1.2 Psychology

Psychology also possesses a long history of investigating phenomena of interest through the use of experimentation. Although there may appear to be less obvious overlap between the topics of interest in psychology and political science than there is between economics and political science, most of the methodological developments used by economists studying individual choice behavior were pioneered by psychologists working in the same area. Therefore, the historical development of psychological experimentation remains methodologically primary. Furthermore, the impact and scope of experimentation in psychology remains far greater than that in behavioral economics.

The scientific study of psychology began late in the 19th century, focusing primarily on the study of human consciousness through processes of introspection. Among the earliest laboratories designed expressly for use in experimentation of psychological phenomena were those built by Wilhelm Wundt in Germany, his student Edward Titchner at Cornell, Edward Thorndike at Columbia, and William James at Harvard. Thorndike (1898) began his laboratory work around the turn of the 20th century by studying learning behavior in cats; this was around the same time that Pavlov developed his theory of classical conditioning by inducing his famous Russian dogs to salivate in response to a bell that originally rang in conjunction with feeding. B. F. Skinner (1966) later built on these behavioral experimental beginnings using mostly pigeons to construct his systematic program of research, which he labeled the *experimental analysis of behavior.*

The history of social psychology informs some of the differences between experiments in behavioral economics and those in social psychology. The most important of these differences lies in attitudes and practices toward the use of deception in experiments. In economics, it is virtually impossible to get an experiment involving deception published in a respectable journal. This is because deception is understood to be antithetical to the purpose of most economics experiments, which are designed to involve transparent instructions with clear payoffs. In social psychology, however, more than half of the studies published in the two most prominent journals (*Journal of Personality and Social Psychology* and *Journal of Abnormal Social Psychology*) involved the use of some sort of deception (Adair et al. 1985). Although this last study dates from the mid-1980s, there has been no change in the use of deception in experimental social psychology since about that time.

In psychology, the history and goals of the discipline have led to different practices. In fact, several of the most important and influential experiments ever conducted involved deception. Social psychologists believe that much of what they know about human behavior has come through the use of deception, at least in part, and that deception can be very useful as long as it is used in an ethically responsible fashion.

Two examples suffice to make the point about the importance of deception in psychological experiments. One of the most famous experiments in social psychology is the Asch (1952) experiment on conformity effects. Subjects in the Asch experiment were told that they were about to participate in an experiment about perception, and were then asked to match the length of one line to one of three other lines that differed in length from one another. As in the familiar Sesame Street games, only one line matched, and the match was perceptually obvious. In this study, Asch demonstrated that most people change their behavior in order to conform to the group, even when this means going against their own perceptions. Obviously, subjects in this experiment might have proved much less likely to conform if they were told that the experiment was about conformity.

Similarly, Milgram's (1974) famous experiment on obedience was billed as a study on learning. Subjects were told that they had to shock a "learner" (in fact, a confederate) when that person failed to answer a paired word association correctly. When the "learner" began to complain about the shocks, mention a heart problem, scream, and then eventually go silent, most subjects became incredibly uncomfortable. Yet most subjects continued to shock their partner up to the maximum amount of shock after the experimenter told the subject that the experiment must continue, and that he would take responsibility. Clearly, subjects' responses might have been quite different if they knew that the experiment was really about obedience and not about learning.

To contrast the central characteristics of experiments in behavioral economics with those in psychology, some overlap exists, but some strong differences emerge. Although economists insist on clear instructions, for example, psychologists often purposefully give

ambiguous directions, especially if they want to study various forms of information processing. As noted, psychologists are more accepting of deception. Both disciplines tend to respect the importance of subject anonymity. However, psychologists are not as prone to offer cash in return for participation in experimentation. For instance, oftentimes students are simply given course credit for their participation. This is because psychologists often feel that monetary rewards impose an artificial reward system that might skew the processes they seek to investigate. Natural social processes, for example, might be hidden by cash incentives. Finally, although behavioral economists often use experiments in order to test formal models, psychologists are more often interested in less mathematical, if no less rigorous, theories of human behavior. In psychology, experiments stand on their own more often than they might in experimental economics, where they are often used in conjunction with more formal models.

Some of the different emphases on deception between economics and psychology derive from the different natures of the questions they investigate. Although experimental economists focus on the topics mentioned previously, social psychologists study such topics as person perception; stereotyping and prejudice; social identity; attitudes, attitude change, and behavior; group dynamics; relationships; and decision making. Some of these topics would be impossible to investigate accurately without the use of deception. For example, topics such as racial prejudice would be difficult for a psychologist to study explicitly because many subjects might lie about their true attitudes and beliefs in order to present themselves as more acceptable to the experimenter. Note that an economist would be very unlikely to be interested in prejudice as a research topic, and therefore less likely to need deception for such purposes. Thus, some of the differences between economics and psychology surrounding the use of deception, as well as other things, derive from the different histories and cultures of experimentation in each discipline.

1.1.3 Political Science

Perhaps the first foray into experimental work in political science was published by Harold Gosnell in the *American Political Science Review* in 1926. This "experiment" on an attempt to stimulate voting really presented what would now be called a field experiment or a field study, discussed at greater length below. With a few exceptions in the early 1950s, more experimental work in political science waited for the pioneering contributions of William Riker (1967), Riker and Zavonia (1970), and Charles Plott (1967, 1979). What is interesting about much of the early work that used experiments to investigate political phenomena is that it appeared in economics and psychology journals much more than in political science journals. In fact, because many believed that their experimental work had been unfairly rejected from the established political science journals (McConahay 1973), the *Experimental Study of Politics*, a new journal expressly founded to redress this difficulty, was established in February 1971. The broader field was not able to sustain enough interest, and the journal lasted a little over four years before being discontinued.

Currently, experimentation is not a major methodology used by political scientists. A comprehensive overview of experiments published in mainstream political science journals reveals a total of 105 articles since 1926. About 80% of these articles have been published in one of five journals, and most of the experiments relate to American voting behavior. The second most popular topic concerned bargaining (13), with games (10) and international relations (10) tying for the third most popular topic. Other topics that attracted experimental interest included committee work (8), experimental bias (6), race (6), field experiments (5), media (4), leadership (4), and experiments embedded in surveys (3).

It is clear that many of these topics in political science overlap with those investigated by behavioral economists and social psychologists. Although the venues and contexts for investigation may differ, a great potential for substantive collaboration exists and might be exploited for greater cross-disciplinary and interdisciplinary work among political science, psychology, and economics. Current areas of research in psychology and economics that offer particularly promising opportunities for collaboration with political scientists include social preferences, including investigations of norms, social networks, altruism, status and trust; bounded rationality, involving decision making in complex environments; learning and expectation formation; attitudes toward risk; and cognitive biases (Laibson 2000).

However, significant differences between the disciplines also exist. Psychology and behavioral economics share some core assumptions about human nature and experimentation that many political scientists do not: a deep skepticism toward notions of rationality in human behavior; emphasis on experimental validation of modeling assumptions; integration of micro level data; and adoption of the lessons about human cognitive processing, from experimental social and cognitive psychology to models of human behavior (Laibson 2000).

2 Methodological Issues in Experimentation[2]

It may prove difficult for psychologists or behavioral economists to work easily with political scientists who do not share some of these same starting fundamental assumptions about human nature. However, some divides are the result of a lack of training and experience and a failure by political scientists to fully understand the methodology of experimentation and its interpretation. As some of this confusion abates, more productive conversations and collaborations can be achieved. In that spirit, this section discusses some of the more difficult and contentious aspects of experimentation for all its practitioners: impact versus control; mundane versus experimental realism; internal versus external validity; deception; and laboratory versus field experiments.

2.1 *Impact versus Control*

Aronson et al. (1995) offer a nice summary of the problem of impact versus control in experimentation when they write:

> On the one hand, the experimenter wants the experimental situation to be meaningful and involving to the subject—in a word, the treatments should have impact. On the other hand, a situation which the subject finds meaningful and involving is also likely to trigger a wealth of memories and influences from the past which can affect the subject's interpretation of present circumstances. From the experimenter's point of view, these memories and influences constitute extraneous variability and jeopardize control over the effects of the independent variable. The choice of an empirical realization of one's conceptual variable—and an appropriate stimulus situation in which to apply it—always represents a compromise between impact and control, in which case a little of each is sacrificed. (pp. 44–45)

In other words, an experimenter wants an involved subject, but not one who brings so much extraneous "baggage" to the task that his or her response is overly biased. This can be hard to establish in practice within a restricted experimental protocol.

Probably the most important advantage attributed to experiments is the ability of the investigator to control the experimental situation. Indeed, the entire premise of experiments is built on the notion that everything that can be controlled should be controlled so that the

[2]This entire section draws heavily upon, and greatly benefits from, discussion and commentary on these issues in Aronson et al. (1995).

environment is identical except for the independent variable, which differs across conditions. In this way, any differences in the outcome measure can confidently be attributed to variations in the independent variable.

Yet there are certainly limitations to total control. Subjects come in with a host of prior experiences, beliefs, expectations, and skills that might have an impact on the outcome of any given experiment, regardless of the intended manipulation. Most of the time, experimenters hope that the random assignment of subjects to conditions eliminates any systematic source of bias by ensuring that it is randomly distributed, and does not systematically effect one condition or another disproportionately. However, total control of a given experimental situation might not always be attainable, and it is possible that certain extraneous factors might still influence the outcomes. This can happen, for example, if the random error serves to drown the experimental effects, or if the independent variable is correlated with another unintended and unidentified aspect of the situation or subject. Sometimes running more subjects can ameliorate this problem, but sometimes it cannot.

However, lack of experimental control might occur precisely *because* of the way subjects interact with the experimental manipulation. Obviously, an experiment of an hour or so cannot erase a lifetime of experiences and attitudes. Subjects bring very different baggage to a given experiment and different individuals can react to the same experimental situation quite differently; they can literally see and experience a different experiment than each other because of the cues the experiment provides to events in a subject's past. Other limitations to an experimental situation arise as a result of ethical constraints, time limitations, and experimental biases such as demand characteristics. By and large, the experimenter is aware of these concerns and does the best job possible to achieve as much experimental control as is feasible.

The conflict arises when an experimenter must create a situation that is sufficiently involving and engaging to subjects to generate real attention and impact, and must do so generally in under an hour. This can often be difficult to do in a practical way. Economists often try to get around this problem by offering money, assuming that this incentive alone guarantees engagement. Often, this works, but those who have performed experiments in certain environments in which the average subject is much wealthier than the experimenter understand the futility of this option. Also, as noted, money alone does not eliminate other sources of bias, particularly social desirability biases. Furthermore, for psychologists studying topics that involve internal sources of motivation, money in fact detracts from investigating the topic of interest by providing an extrinsic reward that may serve only to lower intrinsic motivation for the task at hand (Lepper and Green 1978).

The goal of clever experimentation is to increase the realism and interest of the task without sacrificing control. In theory, it should be possible to maximize both goals simultaneously. In practice, increasing impact often means losing some element of control. Those who do field experiments, for instance, lose a lot of control in order to study real-life situations. However, many paper and pencil surveys in the classroom that possess a high degree of control fail to engage the subjects in even the most minimal ways. Subject boredom can render an experiment as fatal as any other loss of control.

2.2 *Mundane versus Experimental Realism*

Some skeptics' concern about experimentation derives from a misunderstanding of the difference between mundane and experimental realism in experiments. However, when critics argue that an experimental situation is not "realistic" because it does not look like a real-world situation on the surface, they are failing to grasp that in experimental settings, the

subjects' engagement with the underlying process is much more important for experimental control, and thus internal validity, than the external trappings of the situation.

Experimental realism is crucial to experimental success; mundane realism is not. Experimental realism refers to impact in its most important sense: Do subjects believe the situation, problem, or issue they confront? Does it engage and interest them? Does it capture their attention? If so, then experimental realism has been achieved; if not, then it does not matter how much the trappings of the environment resemble a real-life situation, indeed, it does not matter if the experiment takes place in real life, like a field experiment, because the experiment failed to achieve the experimental realism necessary to obtain accurate results.

Mundane realism refers to the similarity between the experimental situation and the real world. Yet just because a situation appears to look like a real-world event or problem on the surface does not mean that it elicits any meaningful behavior worthy of investigation or consideration. One could conduct an experiment on individuals sitting on a couch watching sporting events on television, eating chips and drinking beer, and it would have a great deal of mundane realism because many people spend countless hours engaging in this real-life behavior. However, this study may not engage subjects in any meaningful phenomena of interest because it does not possess experimental realism.

Iyengar (2000) tells a great story about how when he first set up an experimental living room in a local mall complete with couch and television in order to study the effect of campaign ads, subjects did what they do at home: they picked up the remote control and changed the channel when the ads came on. This experiment contained so much mundane realism that the experimental manipulation was avoided altogether. Once the remote was removed from the situation, making it less "realistic" in the mundane sense, the experiment became much more realistic in the experimental sense because subjects now were forced to sit through watching the ads of primary interest to experimenters. This slight reduction in mundane realism in fact allowed and enhanced the more important experimental realism in the study.

Obviously, as the Iyengar study illustrates, experiments can possess both high levels of mundane and experimental realism. Yet many critics assume that because laboratory experiments take place in stripped-down, stylized settings, they cannot tell us anything meaningful about the real world, and this is a mistake. One of the reasons that experimentalists try to achieve high levels of both impact and control in their studies is because of the need to overcome the preexisting beliefs, memories, expectations, and experiences with which subjects enter a study. Experimenters have only a brief period of time to engage the subject in an experience that is sufficiently interesting and meaningful that it overrides these potentially confounding extraneous biases. However, experiments that involve subjects in a meaningful process should be able to overcome the seeming sterility of the environment. As subjects become sufficiently interested in the task or process, their focus remains on the manipulation itself, and not on distractions in the environment that may merely serve to arouse or cue triggers to unrelated processes or events from their past.

One additional point should be considered in assessing the value of experiments for real-world events. Real-world events would be interpreted differently if these events took place in the laboratory, just as laboratory events that take place in the real world appear differently as well. In this way, the criticism of external validity goes both ways. Experiments that are high in mundane realism do not necessarily increase the generalizability of the findings if they do not achieve internal validity and control over the variables of interest first. Rather, experiments in the laboratory that prove meaningful for subjects need not take place in a real-world situation to provide valid results. The laboratory allows for control over potential biases that might interfere with the process in the real world, and allows investigators to

isolate causation, test theories, and generate new hypotheses about human behavior. Once these pieces are isolated and examined, additional complicating factors can be added under some kind of controlled order to begin establishing how these forces might interact in the real world. In this way, the microfoundational basis of larger processes can be uncovered. Then, slowly and through replication, similar studies can be aggregated into a more comprehensive understanding of broader topics and issues.

2.3 *Internal versus External Validity*

External validity is another area that tends to preoccupy critics of experiments. This near obsession remains particularly problematic because issues of internal validity are of much greater concern to any sophisticated experimentalist than are questions of external validity. This causes at least two important points of confusion. First, questions surrounding external validity tend to be used to dismiss experiments that in fact provide a valid basis for generating hypotheses, testing theories, and building models. Second, this confusion results in many critics of experiments being more willing to accept deeply flawed experiments. Some experiments have the surface appearance of external validity because they appear high in mundane realism. The seemingly obvious generalizability often rests on a foundation wholly without merit because the actual experiment offers no internal validity. In short, without internal validity, there can be no external validity.

Prior to concerns about internal or external validity, experimenters must first be careful that their measures possess construct validity. Both the independent and dependent variables should be measured in ways that correspond to the *theoretical* concepts under investigation. At least two problems arise with this concern. First, experimenters may use manipulations that inadvertently affect other concepts, and this introduces measurement error that could bias results in systematic ways. In other words, an experiment might be examining more than is intended in ways that might compromise the findings that are emphasized.

Second, experimenters may use measures that actually do not assess the concepts being investigated. Demand characteristics and social desirability responses within the experiment itself can exacerbate these problems. In other words, an experimenter might not actually be studying the phenomenon of central interest, but rather something else entirely. The best strategies for avoiding these concerns are to use the best and most precise manipulations and measurements possible, and to use *multiple* manipulations and measures wherever possible. This latter issue is also important in assessing external validity (Smith and Mackie 1995).

Internal validity refers to whether you are studying what you think you are studying. Can an experimenter be confident that changes in the independent variable caused changes in the dependent variable? This is determined primarily by the experimental protocol itself, which specifies how subjects are recruited, manipulated, and assessed. The goal is to make all the conditions of an experiment as similar as possible in every way except the manipulation itself. To the extent that this is possible, the experimenter can be confident that the manipulation of the independent variable is the *cause* of any differences seen in the dependent variables. This ability to get at causation is, after all, the main advantage of experimentation, and an advantage offered by no other methodology. Thus internal validity tells us whether we can be confident that the effects are a result of the putative causes. Without high levels of internal validity, it becomes impossible to render accurate causal claims. Internal validity is aided by high levels of control, impact, and experimental realism.

Measuring effects is not the same as manipulating effects because other external factors might be the true cause of a perceived effect. Therefore, voting studies, particularly those that rely on public data on outcomes, cannot confidently attribute changes in voting to field

interventions alone, for example, because too many other factors not under experimental control might have affected the outcomes. For example, a study that examines population demographics and then relates those findings to voting patterns may have established a correlational pattern, but it has not conducted an experiment because there was no random assignments to treatment conditions or manipulations involved in associating the public records data. Obviously, this kind of research is fine; it is just not experimental in nature. Random assignment of subjects to conditions is the crucial best defense against threats to internal validity.

Concerns about internal validity should always take place *prior to* concerns about external validity. To the degree that an experiment is lacking in internal validity, external validity should be a moot point. If we are not sure that *A* caused *B* because of some flaw in the experimental research design, does it really matter whether *B* is representative of, or can be generalized to, the real world? Aronson et al. (1995, p. 75) emphasize the primacy of internal validity when they state, "But of the two, internal validity is, of course, the more important, for if random or systematic error makes it impossible for the experimenter to draw any conclusions from the experiment, the question of the generality of these conclusions never arises."

Regardless, questions concerning external validity continue to elicit much greater concern within political science. This belief merely betrays, inadvertently, a failure to understand the methodological nature of experimentation itself. There are at least two important sources of confusion surrounding this topic. First, many critics seem to think that experimenters are trying to say something about the real world from a single study with college undergraduates. This is almost never the case. Psychologists and behavioral economists recognize that external validity is established over time, across a series of experiments that demonstrate similar phenomena using different populations, manipulations, and measures. External validity occurs through replication. Aronson et al. (1995, p. 82) note that "[b]ringing the research out of the laboratory does not necessarily make it more generalizable or 'true'; it simply makes it different."

Sometimes more limited subject populations or manipulations are sufficient to establish external validity because the issue of concern is limited to that population. This might occur, for example, in studies concerning patient compliance with physician instructions. Studies limited to medical patients are entirely sufficient for investigating this issue because the findings are not intended to be generalized beyond this population. However, in our work on sex differences in aggression, for example, we report the first of a series of experiments that eventually use different populations and different measures to try to further investigate the many issues involved in this topic (McDermott and Cowden 2001). Indeed, it would be irresponsible and inappropriate to draw preliminary conclusions about all world leaders, or all female leaders, for example, from a limited study.

This raises the second important area of confusion. For most experimentalists in psychology or behavioral science, whether a particular study generalizes to a larger audience is really not the most important concern. Rather, experimental studies are intended to test theories and build hypotheses about larger issues involving human behavior. So, for example, I might not expect to find the same behavior in college undergraduates as I would in world leaders, but I would hope that my experimental studies help illuminate more universal properties related to aggression within and across the sexes that might prove more universal in nature. Therefore, the idea of experimentation is *not* primarily that a specific set of experimental results apply directly to broader conclusions about human behavior in other settings, but rather that these results develop and test theories that then, in turn, and in aggregation, help explain and predict certain underlying causal mechanisms in more universal issues (Smith

and Mackie 1995). Thus, experiments should not be evaluated based on appearances of broader generalizability, but rather on the merit of their underlying theoretical constructs.

Aronson et al. (1995, p. 81) summarize the issue as follows:

> Is the ability to *generalize* to other people, places, and times in the real world—to ensure that our findings are "*externally valid*" as the term is traditionally used—necessarily the standard against which the value of experiments . . . should be assessed? Not always. Numerous important experiments . . . have been conducted by researchers whose goals were not to generalize to other subjects, operationalizations, or other settings. And this does not diminish the magnitude of their contributions to understanding behavior, for their goals were not to *make* generalizations, but rather to *test* them.

There are several criticisms that seem to characterize arguments about external validity within experiments. Typically, these relate to the triviality or artificiality of the experimental situation. Indeed, many experimental settings are artificially sterile and unrepresentative of the environments in which subjects might normally perform the task required. This occurs at least partly as a result of the artificial nature of the experiment itself: time is restricted, conditions are controlled, and extraneous variables are eliminated. The artificial nature of the experimental setting may make it difficult for subjects to act as they would in a more natural setting. Yet environments are partly artificial because this also serves as a way to reduce bias and engage subjects in an experimentally, and not mundanely, realistic way.

However, many aspects of real-world complexity are difficult to simulate in the laboratory. Cultural norms, relationships of authority, and the multitask nature of work itself might invalidate any results that emerge from an experiment that does not, or cannot, fully incorporate these features into the environment or manipulation (Walker 1976). In particular, subjects may behave one way in the relative freedom of an experiment, where there are no countervailing pressures acting on them, but quite another when acting within the constrained organizational or bureaucratic environments that operate at their political jobs. Material and professional incentives can easily and quickly override more natural psychological or ethical concerns in the real world. Yet the same forces can manifest themselves more readily in the unconstrained environment of the laboratory. Failure to mimic or incorporate these constraints into experiments, and difficulty in making these constraints realistic, might restrict the applicability of experimental results to the real political world.

However, much of this artificiality is purposeful because it helps isolate the variables of interest. As noted, it is more important to create a research design high in experimental realism than to mimic every aspect of a real-world situation. It is, in fact, experimentally useless to conduct such mundane mirroring. Oftentimes arguments concerning triviality or artificiality are based on the false assumption that similar variables should behave in identical ways in real-world and laboratory settings. Clearly, however, this is not the case because so many other factors besides those variables that can be isolated in the laboratory influence real-life events. Any experimenter interested in a particular human behavior or process must first isolate and study various aspects and features of that behavior under controlled conditions before being able to extrapolate or expand those findings into broader contexts. This control allows for the crucial element of internal validity that establishes the viability of causal claims. Layers of complexity and interaction can thus be added over time in a sequence of experiments in order to reach higher levels of theoretical development and external validity.

2.4 *Deception*

Issues concerning deception have been mentioned in terms of their differing histories in behavioral economics and social psychology. As previously mentioned, deception is

essentially never used in experiments in experimental economics, at least partly because it is not necessary in order to study the phenomena of interest to economists, which are mostly transparent economic transactions.

Yet many topics that psychologists and, potentially, political scientists may be interested in studying might be impossible to investigate accurately without the use of deception. The purpose of deception is simply to keep the subjects from being aware of the actual focus of study until after the experiment is over. Typically, this is because the experimenter believes that if subjects were aware of the true nature of the study, they would change their behavior in ways that would bias the results of the experiment in misleading ways. Often deception is elaborate, as in the case of the Milgram experiment, but at other times deception exists as a sin of omission, whereby the experimenter simply fails to inform the subject of the real purpose of the experiment.

Many observers tend to find deception distasteful almost on principle. However, there is no empirical evidence that experiments involving deception are more harmful to subjects than experiments that do not involve deception (Christensen 1988). Furthermore, subjects who participate in experiments involving deception report enjoying the experience more, and learning more from it, than subjects who participated in nondeceptive experiments (Smith and Richardson 1983). As Aronson et al. (1995) note, this may simply be because experiments involving deception may be more engaging and interesting to subjects than more boring, albeit transparent paper-and-pencil survey-based experiments. In addition, subjects appear to resist telling their friends about the true nature of an experiment after debriefing when asked to do so (Aronson 1966).

The main concerns surrounding the use of deception have to do with invading a subject's privacy, but there are practical methodological concerns with deception as well. In general, subjects rarely correctly discern the true purpose of an experiment, unless the so-called deception is in reality pretty transparent (Brock and Becker 1966; Fillenbaum 1966; Striker et al. 1969; Kruglanski 1975). However, the more deception is used by experimenters, the more suspicious the subject pool as a whole can become, and the more likely an individual subject might be to guess the real purpose of the experiment. Research shows that general suspicion on the part of subjects does increase the variability of results (Kelman 1968; Striker et al. 1969). In addition, participation in an experiment involving deception increases subjects' desire to present themselves as desirable in the next experiment. It also makes subjects less compliant with subsequent experimental demands (Silverman et al. 1970). Even the mere distraction of subjects trying to discern the real reason for the experiment might bias the results in a systematic way. Certainly the worst subjects in this regard are those who participate repeatedly in experiments for money; such subjects become experienced and can become more suspicious and sophisticated about hypotheses. Although economic experiments demand monetary compensation to ensure the rational behavior of subjects, psychologists have reason to believe that paying subjects serves only to undermine crucial naivete in certain populations.

2.5 *Laboratory versus Field Experiments*

As noted, the earliest experiment conducted in political science was a field experiment by Gosnell. Some practitioners continue to have more enthusiasm for field experiments than psychologists, mostly because, as noted previously, they tend to be more concerned with mundane as opposed to experimental realism, and questions of external as opposed to internal validity, although this is methodologically backward.

Field experiments definitely have their place in experimental research, however. First, field experiments can be the only alternative available in some situations in which the experimenter either cannot or should not create the desired situation within a laboratory setting. Sometimes this is because field experiments are the only way to investigate certain events, such as the impact of a natural disaster such as an earthquake or tornado; no laboratory simulation could ever come close to the impact and realism of the real thing. Indeed, it would be unethical to create such conditions experimentally; an experimenter would not be able to study the effects of a life-threatening illness by causing disease in a subject, for example. Second, it may be very hard to simulate accurately many all-encompassing phenomena of interest (e.g., an election, a war, an economic recession). Skilled and clever experimentalists excel at being able to distill the basic processes under investigation into an experimental manipulation, but the task is neither easy nor straightforward.

However, a field experiment is not, and should not be, simply a laboratory experiment conducted in the real world. In general, field experimenters must sacrifice a great deal of experimental control in order to gain some advantage in either mundane realism or external validity. This lack of control requires that many important accommodations need to take place in field experiments in order to garner any findings that have the necessary internal validity to assess causality confidently.

First, it is not enough simply to measure preexisting phenomena and call that an experiment because certain factors, such as demographics, correlate with other factors, such as voting patterns. Second, researchers must be acutely sensitive to the many extraneous factors over which they retain no control, but that can have an impact on their findings in the field. Third, and perhaps most important, it is often impossible to assign subjects randomly to conditions in the field. Sometimes this is ethically impossible, as it would be to study the occurrence of deadly illness by infecting healthy people randomly. Yet often such failure serves to limit the very generalizability of results in ways that experimenters may fail to recognize or acknowledge.

Effective laboratory situations strive to be coherent, simple, and engaging. Control over the random assignment of subjects to conditions ensures that all subjects start from the same baseline as much as possible (Aronson et al. 1995). Similarly, experimenters in the field try to choose a situation that exists naturally but that mirrors as closely as possible the controlled conditions that can be created in a laboratory. Obviously this is harder to achieve in the field where there are many more extraneous factors that might affect the recruitment, manipulation, or measurement of subjects and outcomes. In evaluating the findings reported by field studies, readers must be careful to assess the internal validity of the study before buying into the generalizability of the results.

Finally, experimenters in the field face more challenging ethical concerns because oftentimes their subjects are unaware that they are being studied, and have not given informed consent to be investigated. Although this may aid in achieving subject nonreactivity, lack of awareness of participation in an experiment constitutes a serious invasion of subject privacy. Under such circumstances, experimenters are especially obliged to impose as little stress as possible on subjects, and to cease immediately any experimental procedure that appears to cause any harm to subjects.

3 Advantages and Disadvantages of Experiments

Because most readers are well aware of the threats and potential responses to issues surrounding internal and external validity, and, now the central issues of concern to experimentalists, how can we evaluate the relative merits of experimental methodology as

opposed to other research strategies, whether they be case work or formal modeling? As an advocate for experimentation, I place more emphasis below on advantages over disadvantages. Part of my argument is that experimentation is underused in political science and might work to the advantage of practitioners who are seeking a more definitive form of causal inquiry, a useful methodology for the cumulation of knowledge, or an additional tool in their methodological arsenal for triangulating on problems that appear intractable when using other forms of methodological inquiry.

3.1 *Advantages*

Many of the issues discussed in this section overlap with some of the previous discussion, but should be mentioned independently for their methodological advantages. There are at least four such advantages that experiments offer: (1) the ability to derive causal inferences; (2) experimental control; (3) precise measurement; and (4) the ability to explore the details of given processes (McConahay 1973).

The primary comparative advantage of experiments lies in their potentially high degree of internal validity. Correlational studies, for example, obviously do not establish causation. Because a laboratory setting allows investigators to control all aspects of the environment so that only the independent variables differ, any differences in the dependent variable can be attributed to the manipulation, and thus offer support for causal inferences. No other methodology can offer such strong support for the causal inferences that experiments allow. Thus, "the major advantage of laboratory experiments is in its ability to provide us with unambiguous evidence about causation" (Aronson and Carlsmith 1968, p. 10). Experiments can prove quite useful in helping to develop and test theories to explore whether hypothesized relationships hold and under what conditions they are operative.

Second, experiments allow the investigator to have unparalleled control over the testing environment. Experimenters not only create the experimental protocol and control the experimental environment, but most importantly, they also control the random assignment of subjects to conditions. This randomization process creates the pretreatment similarity of the groups, so that observed differences in outcome can be attributed confidently only to manipulations, which take place within the experiment itself, and not to preexisting differences within the subject pool. Moreover, the experimenter can decide which independent variables to manipulate, and can vary the timing and intensity of those variables at will. This process allows the examiner to test for alternative explanations for the hypothesized relationships as well.

Third, laboratory experiments allow for the precise measurement of the variables of interest. Experimenters need not overly concern themselves with improper measurement because they design and implement the desired measures themselves. This allows experimenters to eliminate extraneous factors that might contaminate a study by inducing spurious results. Particularly with the increase in the use of computers to run experiments, even more precision is possible because potential errors in coding, transcription, or data entry are substantially reduced as well.

Fourth, experiments offer the opportunity to explore the phenomenon of interest in great detail. Complex relationships can be broken down, isolated, and investigated in smaller units in order to see which part of the process results in the phenomenon of interest. In addition, experiments allow particular relationships to be explored in the presence or absence of other variables, so that the conditions under which certain relationships hold can be examined as well. In this way, process can be studied as well as outcome.

Finally, another advantage of experiments results from the scientific rigor designed into the process. Experimenters must be aware of, and control, the independent variables of interest. They must be careful in recording results. Statistical analysis allows for the detailed testing of the relationships between these variables and any interactions among them. With this process, results and insights that might not have been obvious to less systematic or larger scale analysis become prominent. Experiments allow causal inference, control of variables, precise measurement, and clarity of detail.

3.2 *Disadvantages*

Clearly, however, experiments are not a panacea for all methodological concerns. Experiments are not the only, or even the best, form of investigation for all research questions. Indeed, experiments are only one of many methods that can be used to examine political phenomena of interest. As noted, most of the popular concerns about the disadvantages of experiments revolve around questions of external validity and how widely the findings of the laboratory apply to real-world actors and phenomena. Many of these concerns about experimentation, including the artificial nature of the laboratory, the unrepresentative nature of the subject pool, and questions surrounding external validity, have been discussed in detail above.

From the perspective of an experimenter, the main concern with experiments revolves around problems posed by experimental bias of one sort or another. Clearly, an experiment must consider alternative explanations for whatever results are discovered, but experimenters must also be concerned about the confounding variables they bring to the experimental situation that might affect certain findings. The most important of these artifacts are experimenter bias and demand characteristics.

Experiments can be biased by an experimenter setting up an experiment in such a way as to validate prior expectation and beliefs, often unconsciously. Experimenters or research assistants who are privy to the experimental hypotheses might induce subjects to react in ways that confirm these expectations without even meaning to do so (Rosenthal 1966, 1969). Such experimenter bias presents a potentially debilitating confound that can threaten the internal validity of a study.

Subjects themselves also tend to be exquisitely sensitive to implicit experimenter demands and expectations. Note that this is different from subjects being able to discern the relevant hypotheses being investigated in the study. Such demand characteristics and expectancy effects threaten the construct validity of the experiment (Orne 1962; Smith and Mackie 1995). In addition, subjects want to put themselves in a good light, creating a very strong social desirability bias (Rosenberg 1966, 1969). There is also the infamous "Hawthorne effect," whereby subjects change their behavior based solely on the recognition that they are being observed, which can limit the relevance, generalizability, or accuracy of certain experimental results as well (Roethlisberger and Dickson 1939).

The best antidote to these concerns, if possible, is to keep those who run the experiment, or who code the data, blind to the condition of each subject and to the experimental hypotheses. In this regard, the use of electronic collection, such as afforded by computers, can be of immense help in reducing experimental bias.

4 Conclusions

Experiments have a long and distinguished history of effective use in other disciplines, including medicine and psychology. They have recently acquired great cachet in the field of experimental economics as well. They have been slower to acquire a following of

practitioners in political science, mostly due to concerns surrounding the external validity of these kind of studies.

Experiments offer a clear advantage over other methods in particular areas of investigation. In behavioral economics, they are used to validate theories developed by formal modeling. In psychology, they uncover basic processes in human judgment and decision making, among other things. In political science, they can help in triangulating on a process that has proved intractable, or presented contradictory findings, using methods that allow for less clear causal inference. In all areas, they can support theory development, testing, and refinement. In particular, experiments can offer many useful insights in work that investigates the underlying process of a particular phenomenon as opposed to its outcome. Experimentation may not prove as useful for single case examples, but it can serve a useful purpose, as it has in behavioral economics, to help advance and cumulate knowledge more quickly and carefully.

References

Adair, J., T. Dushenki, and R. Lindsay. 1985. "Ethical Regulations and Their Impact on Research Practice." *American Psychologist* 40:59–72.

Aronson, E., and M. Carlsmith. 1968. "Experimentation in Social Psychology." In *The Handbook of Social Psychology* (rev. ed.), eds. G. Lindzey and E. Aronson. Reading, MA: Addison-Wesley, pp. 1–79.

Aronson, Eliot. 1966. "Avoidance of Inter-Subject Communication." *Psychological Reports* 19:238.

Aronson, Eliot, Pheobe Ellsworth, J. Merrill Carlsmith, and M. H. Gonzales. 1995. *Methods of Research in Social Psychology*, 2nd ed. New York: McGraw-Hill.

Asch, Splomon. 1952. "Effects of Group Pressure upon the Modification and Distortion of Judgments." In *Groups, Leadership, and Men*, ed. H. Guetzkow. Pittsburgh: Carnegie Press, pp. 177–190.

Bernoulli, D. (1738). "Specimin Theoriae Novae De Mensurea Sortis." *Commentarii Academiae Scientiarum Imperialis Petropolitanae* 5:175–192. (English translation in *Econometrica*, 1954, 22:23–36.)

Brock, T., and L. Becker. 1966. "'Debriefing' and Susceptibility to Subsequent Experimental Manipulation." *Journal of Experimental Social Psychology* 2:314–323.

Christensen, L. 1988. "Deception in Psychological Research: When Is Its Use Justified?" *Personality and Social Psychology Bulletin* 14:664–675.

Economist. 2000. "Rethinking Thinking." December 16.

Fillenbaum, S. 1966. "Prior Deception and Subsequent Experimental Performance: The 'Faithful' Subject." *Journal of Personality and Social Psychology* 4:532–537.

Gosnell, H. 1926. "An Experiment in the Stimulation of Voting." *American Political Science Review* 20:869–874.

Iyengar, S. 2000. "*Presentation of Experimental Designs for Political Communication Research: From Shopping Malls to the Internet*." Experimental Methods Conference, Cambridge, MA.

Kagel, Jack, and Alvin Roth. 1995. *Handbook of Experimental Economics*. Princeton, NJ: Princeton University Press.

Kelman, Herbert. 1968. *A Time to Speak Out: On Human Values and Social Research*. San Francisco: Jossey-Bass.

Kruglanski, A. 1975. "The Human Subject in the Psychology Experiment." In *Advances in Experimental Social Psychology*, vol. 8, ed. L. Berkowitz. New York: Academic Press, pp. 101–147.

Laibson, D. 2000. Presentation to the Experimental Methods Conference, Cambridge, MA.

Lepper, Mark, and D. Greene. 1978. "Overjustification Research and Beyond: Toward a Means–End Analysis of Intrinsic and Extrinsic Motivation." In *The Hidden Costs of Reward: New Perspectives on the Psychology of Human Motivation*, eds. M. R. Lepper and D. Greene. Hillsdale, NJ: Erlbaum, pp. 109–148.

Lowenstein, Roger. February 11, 2001. Exuberance Is Rational. *New York Times Magazine*, 68.

McConahay, J. 1973. "Experimental Research." In *Handbook of Political Psychology*, ed. J. N. Knutson. San Francisco: Jossey-Bass, pp. 356–382.

McDermott, Rose, and Jonathan Cowden. 2001. "The Impact of Uncertainty and Sex in a Crisis Simulation Game." *International Interactions* 27:353–380.

Milgram, S. 1974. *Obedience to Authority*. New York: Harper & Row.

Orne, Martin. 1962. "On the Social Psychology of the Psychological Experiment with Particular Reference to Demand Characteristics and Their Implications." *American Psychologist* 17:776–783.

Palfrey, T., ed. 1991. *Laboratory Research in Political Economy*. Ann Arbor: University of Michigan Press.

Plott, C. 1967. "A Notion of Equilibrium and Its Possibility Under Majority Rule." *American Economic Review* 57:787–806.

Plott, C. 1979. "The Application of Laboratory Experimental Methods to Public Choice." In *Collective Decision Making: Applications from Public Choice Theory*, ed. C. S. Russell. Dordrecht, Holland: D. Reidel, pp. 137–160.

Riker, W. 1967. "Bargaining in a Three-Person Game." *American Political Science Review* 61:642–656.

Riker, W., and W. J. Zavonia. 1970. "Rational Behavior in Politics: Evidence from a Three Person Game." *American Political Science Review* 64:48–60.

Roethlisberger, F., and W. Dickson. 1939. *Management and the Worker*. Cambridge, MA: Harvard University Press.

Rosenberg, M. 1966. "Some Limits of Dissonance." In *Cognitive Consistency*, ed. S. Feldman. New York: Academic Press, pp. 135–170.

Rosenberg, M. 1969. "The Conditions and Consequences of Evaluation Apprehension." In *Artifact in Behavioral Research*, eds. R. Rosenthal and R. Rosnow. New York: Academic Press, pp. 249–279.

Rosenthal, R. 1966. *Experimenter Effects in Behavioral Research*. New York: Appleton-Century-Crofts.

Rosenthal, R. 1969. "Interpersonal Expectations: Effects of the Experimenter's Hypothesis." In *Artifact in Behavioral Research*, eds. R. Rosenthal and R. Rosnow. New York: Academic Press, pp. 181–277.

Schelling, T. 1958. "The Strategy of Conflict: Prospectus for the Reorientation of Game Theory." *Journal of Conflict Resolution* 2:203–264.

Siegel, S., and L. Fouraker. 1960. *Bargaining and Group Decision Making: Experiments in Bilateral Monopoly*. New York: McGraw-Hill.

Silverman, I., A. Shulman, and D. Wiesenthal. 1970. "Effects of Deceiving and Debriefing Psychological Subjects on Performance in Later Experiments." *Journal of Personality and Social Psychology* 14:203–212.

Skinner, B. F. 1966. "What Is the Experimental Analysis of Behavior?" *Journal of Experimental Analysis of Behavior* 9:213–218.

Smith, Eliot, and Diano Mackie. 1995. *Social Psychology*. New York: Worth Publishers.

Smith, S., and D. Richardson. 1983. "Amelioration of Deception and Harm in Psychological Research: The Important Role of Debriefing." *Journal of Personality and Social Psychology* 44:1075–1082.

Striker, L., S. Messick, and D. Jackson. 1969. "Evaluating Deception in Psychological Research." *Psychological Bulletin* 71:343–351.

Thorndike, E. 1898. "Animal Intelligence." *Psychology Review of Monograph Supplement* 2 (4, Whole No. 8).

Thurstone, L. L. 1931. "The Indifference Function." *Journal of Social Psychology* 2:139–167.

Uchitelle, Louis. 2001. "Some Economists Call Behavior a Key." *The New York Times*, February 11, section 3, p. 1.

von Neumann, J., and O. Morganstern. 1944. *Theory of Games and Economic Behavior.* Princeton, NJ: Princeton University Press.

Walker, T. 1976. "Microanalytic Approaches to Political Decision-Making." *American Behavioral Science* 20:93–110.

Wallis, W. Allen, and Milton Friedman. 1942. "The Empirical Derivation of Indifference Functions." In *Studies in Mathematical Economics and Econometrics in Memory of Henry Schultz*, eds. O. Lange, F. McIntyre, and T. O. Yntema. Chicago: University of Chicago Press, pp. 175–189.