

EXPERIMENTAL METHODS IN POLITICAL SCIENCE

Rose McDermott

Department of Government, McGraw Hall, Cornell University, Ithaca, New York 14850;
e-mail: rmm21@cornell.edu

Key Words experiments, behavioral economics, political economy, game theory, individual choice, methodology

■ **Abstract** This article reviews the use of experiments in political science. The beginning section offers an overview of experimental design and measures, as well as threats to internal and external validity, and discusses advantages and disadvantages to the use of experimentation. The number and placements of experiments in political science are reviewed. The bulk of the essay is devoted to an examination of what we have learned from experiments in the behavioral economics, political economy, and individual choice literatures.

INTRODUCTION

Anyone who takes an antibiotic, confident that illness will remit, is implicitly trusting in the power and validity of experiments as applied to real-world contexts. Indeed, the hard sciences, including biology, chemistry, physics, and medicine, all rely primarily on experimentation to examine and illuminate basic processes. Psychology embodies a long and distinguished history of experimentation, and behavioral economics, which involves a great deal of experimentation, has recently gained increasing prominence within the larger field of economics. But the methodology of experimentation has been slow to garner a following in political science. Experimentation might easily dovetail with methods more established in political science, such as formal modeling, to produce and cumulate useful knowledge; however, political scientists typically prefer archival work, case studies, field work, surveys, quantitative analysis, and formal modeling instead. Yet these other methods need not compete with experimentation. Indeed, the most exciting opportunity for methodological advancement using experimentation lies at the intersection of formal modeling and experimental testing: Formal models present hypotheses that are tested, refined, and explored through experimentation in a reciprocal manner. This process is widely and successfully employed within behavioral economics. As yet, however, political science remains slow to embrace the added value offered by the methodology of experimentation.

This essay addresses the use of experiments and the experimental method in political science. Following a brief background discussion of the experimental method, including threats to internal and external validity, relative advantages and disadvantages, and ethics, the essay concentrates on what we have learned of substance from experiments in behavioral economics and political science that should be of interest to mainstream political scientists.

My overall goal is to advocate for the utility of experiments for political science. I do not argue that experiments are the only, nor the best, form of methodological inquiry. Rather, I argue that experimentation can be particularly useful under certain circumstances: when existing methods of inquiry have produced inconsistent or contradictory results; when empirical validation of formal models is required; when investigators want to triangulate in on specific processes that have already been examined in a more general way using other methodologies; and when evidence is needed to support strong causal claims. Experiments can combine with other methods to provide what Campbell described as a “fish scale model of omniscience,” whereby each methodological layer serves to illuminate and support other component parts.

EXPERIMENTAL METHODS

What do we mean by experiments? I take the term to refer primarily to laboratory studies in which investigators retain control over the recruitment, assignment to random conditions, treatment, and measurement of subjects. This definition assumes that experimenters take pains to assure that the experimental situation does not vary in any way other than the intended independent variables in order to assure the internal validity that allows causal claims.

To be sure, other forms of experimentation exist, including field studies, field experiments, and even simulation studies. Field studies typically look retrospectively at the effect of naturally occurring events. Although such studies seek to achieve maximum realism and representative subject samples, the experimenter has no control whatsoever over the variables of interest. The ideal subject of a field study is a naturally occurring situation with both “before” and “after” data, such as students’ test results before and after implementation of an educational reform program (Walker 1976).

Field experiments take place outside the laboratory but allow the experimenter to retain some limited control over the central variables. At their best, field experiments can offer a reasonable trade-off between internal and external validity through increased realism without too much loss of control. Simulations offer a similar compromise between the concerns of internal and external validity. Some experiments use people as subjects; other rely solely on computerized models. But for purposes of the discussion below, laboratory experiments are the gold standard.

Experimental Design

Why do we need experiments? We need experiments because they help to reduce the bias that can exist in less rigorous forms of observation. Experiments reduce the impact of bias by introducing standardized procedures, measures, and analyses.

Important aspects of experimental design include standardization, randomization, between-subjects versus within-subject design, and experimental bias.

1. *Standardization* remains crucial in experimentation because it ensures that the same stimuli, procedures, responses, and variables are coded and analyzed. This reduces the likelihood that extraneous factors, of which the experimenter might not even be aware, could influence the results in decisive ways. Standardization requires that the same set of experimental procedures, or experimental protocol, is administered in the same way to subjects across conditions; only the independent variable (or variables) of interest is manipulated. This process ensures that the data elicited from subjects are comparable and are not the result of some extraneous feature in the environment.
2. *Randomization* refers to the assignment of subjects to experimental conditions. Experimenters assign subjects randomly to ensure that no unrelated or spurious factors vary consistently within a given population and therefore bias the results. The idea is that background differences cancel each other out in the course of random assignment, since each individual is as likely to be placed in one condition as in another. Thus, no systematic differences in subjects can bias the results of the study. Many people unfamiliar with probability might think that random assignment is little more than alternately assigning Condition A or B as subjects walk in the door. However, the most elegant and pure way to ensure random assignment of subjects to conditions is to use a random number table to assign manipulations.

Experimental design can be between-subjects or within-subject. Typically, one experimental condition is compared to another experimental condition and then to a control condition. The control condition creates a baseline against which investigators compare the results of their manipulations. In between-subjects designs, different groups of subjects are randomly assigned to various experimental or control conditions. In within-subject designs, otherwise known as the A-B-A experimental design strategy, each person serves as his or her own control by participating across time in both control and treatment conditions. Subjects begin with a baseline (A), are then administered a treatment or manipulation condition (B), and are measured again at baseline (A) once the treatment ends (Zimbardo & Gerrig 1996). The comparison of before and after measures on the variable of concern inform the investigator as to the impact, if any, of the treatment on the subject.

3. *Placebo effects*, in medicine, account for patients whose condition improves as a result of a fake treatment, such as a sugar pill. These effects can cause quite powerful changes in outcome based on an individual's belief that the

treatment will work. Control conditions are important in experimental verification precisely because they help determine the extent of placebo effects in an experimental manipulation.

4. *Experimental bias.* Although experiments seek to maximize experimenter control over the independent variables in a study, the experimental process itself can introduce potential sources of bias. Three important forms of experimental bias are expectancy effects, experimenter bias, and demand characteristics.

EXPECTANCY EFFECTS Expectancy effects occur when an experimenter communicates, usually in a subtle, unconscious way, how he or she wants the subject to behave or respond (Rosenthal 1966). Results then take the form of a self-fulfilling prophecy as experimenters create the reactions they hope to elicit with their subtle signals and not with their controlled manipulation. Ways to overcome this bias include having various experimenters run some subjects in all conditions, making the experimenter blind to the subjects' conditions, designing the experiment to avoid experimenter involvement (as can be done with computer-generated experiments), or treating the experimenter as a factor or variable in the statistical analysis at the end of the experiment to determine if any particular experimenter elicited distinctive responses from the subjects.

EXPERIMENTER BIAS Experimenter bias can overlap with expectancy effects but is in fact distinct from them theoretically. Many experimental choices originate from the experimenter's beliefs and attitudes, and these choices can influence the design of an experiment in a nonrandom way. Sometimes this may be acceptable, but concerns arise especially when an investigator remains unaware that his beliefs have unduly affected the design of the study (Roth 1988).

DEMAND CHARACTERISTICS Demand characteristics are similar to expectancy biases as well, although the cues emerge from the subject's interpretation of the experiment rather than from anything the experimenter says or does directly (McConahay 1973). A systematic bias can be introduced if the purpose of the experiment is too obvious. This bias can be exacerbated if subjects, experiencing evaluation apprehension (Rosenberg 1965), try to make the experimenter like them by doing what they believe the experimenter wants. Ways to reduce the impact of this problem on experimental outcomes include using deception to ensure that subjects cannot determine the relevant hypotheses or tests, evaluating the demand characteristics of the experiment in the analysis at the end of the study, and using computer technology to complicate and/or depersonalize the experiment to reduce the likelihood that subjects can discern its true purpose.

Experimental Measures

Experimental measures strive for reliability as well as internal and external validity. Reliability and validity are central concepts in all experimental measurement.

Reliability refers to the extent to which an experimenter tests the same thing time and time again. A reliable result is one that is easily replicable. Reliability improves when measures are standardized, when a larger number of measures have been taken, and when factors that might bias the data are controlled in advance (Zimbardo & Gerrig 1996).

Experimental measures can take several forms: self-reports, behavioral measures, physiological measures, and incentives.

SELF-REPORTS, BEHAVIORAL MEASURES, AND PHYSIOLOGICAL MEASURES Self-reports are usually verbal or written reports of a subject's responses to a particular set of stimuli. They can take the form of paper-and-pencil questionnaires, surveys, or interviews. Some of these data, which might be originally obtained in qualitative form, can be coded into quantitative categories for later analysis. Behavioral measures require experimenters to observe the behavior of subjects by, for example, videotaping them and later examining the tapes for characteristics such as facial expressions or tendency to dominate in a group. Physiological measures include such data as heart rate, galvanic skin response, blood pressure, or more extensive tests such as magnetic resonance imaging (MRI) or positron-emission tomography (PET) tests. More intrusive tests, such as those that analyze saliva, urine, or blood, might also be conducted to determine hormone levels or other variables of interest.

INCENTIVES Most experiments in psychology do not offer incentives other than extra credit in a course for participants. Indeed, many introductory psychology classes require students to participate in experiments. However, many behavioral economics experiments do offer material incentives, typically money or a lottery that offers the chance of a cash reward to subjects. There are two kinds of incentives: those offered to subjects merely for showing up to participate in a study, and those intrinsic to the experiment itself.

Threats to Internal and External Validity

Campbell & Stanley (1963) describe the classic distinction between internal and external validity as follows: Internal validity asks, "Did in fact the experimental treatments make a difference in this specific experimental instance?" External validity investigates the generalizability of the results by inquiring, "To what populations, settings, treatment variables, and measurement variables can this effect be generalized?" As McConahay (1973) noted, "Generally, with a large number of specific exceptions, psychologists are more concerned about the internal validity of a research design while political scientists are more concerned with external validity." Walker (1976), bemoaning the fact, went so far as to argue that herein lay one of the main reasons for the lack of productive communication between the two disciplines: "The social psychologist, with his reliance on controlled laboratory experiments, has different interests and concerns than the political scientist, who is more at home in the field."

Once an investigator has determined what to investigate and how to assess and measure the relevant variables, he or she must ensure that the design of the experimental protocol does not fall prey to one of the flaws that can confound the results (see Campbell & Stanley 1963 and Campbell & Ross 1968 for the classic discussion of these confounds).

THREATS TO INTERNAL VALIDITY There are nine potential threats to internal validity in experimentation:

1. *History* refers to any event that occurs outside the experimenter's control in the time between the measures on the dependent variable. This phenomenon becomes a concern when there is a lot of time between the measurements on the dependent variable.
2. *Intersession history* refers to events that occur inside the study itself, which are beyond the control of the investigator and may affect the outcome of the study. Extreme temperature fluctuation, unexpected fire drills, or unknown preexisting relationships between some subjects might affect one session of an experiment but not another. These confounds all threaten the internal validity of the experiment.
3. *Maturation* refers to the natural needs, growth, and development of individuals over time. For example, an experiment that relies on deprivation for motivation depends on maturation effects for thirst to occur. Maturation processes that work independently of the investigator over the course of an experiment can bias results as well.
4. *Performance effects*. Performance can change as a result of experience. Test performance can be affected by the very act of having taken the test before. Therefore, pre- and post-tests with exactly the same questions do not constitute identical assessments because, independent of the intervening manipulation, taking the first test may influence the answers to the second test through the natural process of learning and experience.
5. *Regression toward the mean*. Since all scores represent some combination of the real score and some random error, subjects who manifest an extreme score are likely to move closer to the mean on the next measurement, as the random error fluctuates. Experimenters who specifically pick subjects because they manifest an extreme score on some dimension, like authoritarianism, are likely to confound their results through their failure to incorporate regression effects into their subject selection procedures.
6. *Subject self-selection*. Subjects who self-select into particular experiments or conditions are likely to differ in some systematic way from those who are randomly assigned to a condition.
7. *Mortality* occurs not only when subjects die but also when they are lost to follow-up by the investigator. In political science, mortality occurs most often in experiments that require the same subject to show up more than once, and

the subject fails to show up for the second part of the experiment. This can also happen in longitudinal and field studies that examine dynamics over time; as people move and change their living situations, they can become hard to trace. Sometimes financial and other incentives can alleviate this problem. From an ethical perspective, the most important cases of experimental mortality occur when subjects leave in the middle of an experiment because some aspect of the experiment has made them uncomfortable. If a study has a high degree of interexperimental drop-out, the investigator should take pains to ascertain that the experiment is being conducted in an appropriate and ethical manner.

8. *Selection-maturation interaction* occurs when subjects are placed into an experimental condition in a nonrandom manner and some aspect of the group differs in maturation from the other groups in an important or systematic way.
9. *Unreliable measures*. If measures are unreliable, if the subject population shifts, or if some other aspect of the experimental condition is affected in a nonrandom manner, results can become biased.

THREATS TO EXTERNAL VALIDITY Campbell (1968) outlines the six major threats to external validity in experimentation:

1. *Testing interaction effects*. Testing can increase subjects' sensitivity to the variables under investigation, which makes it difficult to generalize the results to a population that has not been pretested.
2. *Unrepresentative subject population*. What can college sophomores tell us about real-world decision makers? Sears (1986) argues that college sophomores differ in systematic and marked ways from other people: They are more self-absorbed; they have less crystallized attitudes, a less clear sense of self, higher rates of compliance, less stable peer relationships, and stronger cognitive skills. Remarkably, many experimental findings using college sophomores have proved remarkably robust (Roth 1988). However, many people remain concerned about subject pools. Obviously, the best way to deal with this problem of external validity would be to sample directly from the populations of interest. Etheredge (1978) did this in his extensive study of 126 career foreign service officers at the State Department to examine "how emotional predispositions might shape elite foreign policy thinking." But often this is not possible because such people are either too busy or not interested in participating in experiments. Another strategy against this limitation involves simulations with real or former decision makers using a hypothetical or past crisis as a stimulus. Such simulations have produced very accurate results. In one of the most powerful examples of a simulation's prescient prediction of a real-world outcome, a Joint War Games Agency of the Joint Chiefs of Staff conducted a major war game simulation of the conflict in Vietnam. The goal was to start with current resources as of July of 1965 and simulate the likely outcome through September of 1966. The results indicated that the United States would not be able to win the conflict in the long

run and was unlikely to do better than stalemate in the short run (as reported in Burke & Greenstein 1989). Simulations that accurately mimic real-life problems and resources can engage the same psychological processes that operate in the real world.

3. *Hawthorne effect*. The third limitation on external validity involves the so-called Hawthorne effect (Roethlisberger & Dickson 1939), whereby people change their behavior merely because they are aware of being observed. People who know they are in an experiment may behave differently than they would if they were not in an experiment or were unaware of the experiment.
4. *Professional subjects*. On large and relatively anonymous college campuses, a student eager to earn money can participate in many experiments across departments without any one experimenter realizing how experienced this subject has become. Overly experienced or jaded subjects may be more likely to guess the underlying hypothesis or manipulation in an experiment if they have participated in similar ones in the past.
5. *Spurious measures*. Some unexpected aspect of the experiment may induce subjects to give systematically irrelevant responses to particular measures, which are then understood to be experimental effects.
6. *Irrelevant measures*. Irrelevant aspects of the experimental condition might produce results that appear to be experimental effects.

ADVANTAGES AND DISADVANTAGES OF EXPERIMENTS

Since most of the threats to internal and external validity can be anticipated, what are the benefits that a skilled experimenter might accrue from using experimentation as opposed to other research strategies, such as case study work or formal models? What problems would still plague the interpretation and utility of his or her results?

Advantages

The comparative advantages of experiments lie in their high degree of internal validity. No other methodology can offer the strong support for the causal inferences that experiments allow. Correlational studies, for example, do not show causation. Since a laboratory setting allows investigators to control all aspects of the environment so that only the independent variables differ, any differences on the dependent variable can be attributed to the manipulation, and thus offer support for causal inferences. Experiments offer at least five such advantages:

1. *Ability to derive causal inferences*. “The major advantage of laboratory experiments is in its [sic] ability to provide us with unambiguous evidence about causation” (Aronson & Carlsmith 1968). Because of the randomization of subjects and the control of the environment, experiments allow

confidence regarding causal inferences about relationships among the variables of interest.

2. *Experimental control.* The experimenter has control over the recruitment, treatment, and measurement of subjects and variables.
3. *Precise measurement.* Experimenters design and implement the desired measures and ensure that they are administered consistently.
4. *Ability to explore the details of process.* Experiments offer the opportunity to explore phenomena of interest in great detail. Complex relationships can be broken down and investigated in smaller units in order to see which part of the process results in the differences 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.
5. *Relative economy.* Although experiments may be more costly and time-consuming than some research methodologies, such as formal modeling, they are certainly more economical than conducting large surveys or field experiments. Students are a relatively inexpensive and reliable subject pool, and a large number of them can be run in a semester. Experiments embedded in larger surveys (Sniderman et al. 1991, Kuklinski et al. 1997) may provide a more representative sample but would also require a great deal more time and money to administer.

Disadvantages

Experiments are not always the ideal methodology. Most concerns about their disadvantages within political science revolve around questions of external validity and how widely the findings of the laboratory apply to real-world actors and phenomena. There are four main disadvantages to the use of experiments:

1. *Artificial environment.* Many experimental settings are artificially sterile and unrepresentative of the environments in which subjects might normally perform the behavior under study. There are at least two important aspects of this limitation. First, it might be impossible or unethical to create the desired situation within a laboratory. An experimenter could not study the effects of a life-threatening illness by causing such disease in a subject. Second, it may be very hard to simulate many phenomena of interest—an election, a war, an economic recession, and so on.
2. *Unrepresentative subject pools.* As noted above, subject pools may be unrepresentative of the populations of interest.
3. *External validity.* For political scientists, questions surrounding external validity pose the greatest concern with experimentation. What can experiments tell us about real-world political phenomena? Beyond the nature of the subject pool, this concern is at least twofold. First, in the laboratory it is difficult to replicate key conditions that operate on political actors in the real world.

Subjects typically meet only for a short period and focus on a limited task. Even when money serves as a material incentive, subject engagement may be low. In the real world, actors have histories and shadows of the future with each other, they interact around many complex issues over long periods, and they have genuine strategic and material interests, goals, and incentives at stake. Can the results of a single-session experiment tell us anything about such a complicated world?

Second, and related, many aspects of real-world complexity are difficult to simulate in the laboratory. Cultural norms, relationships of authority, and the multitask nature of the 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 in which they work at their political jobs. Material and professional incentives can easily override more natural psychological or ethical concerns that might 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.

There are two important things to understand about external validity. First, external validity is only fully established through replication. Experiments testing the same model should be conducted on multiple populations using multiple methods in order to determine the external validity of any given experimental paradigm. Second, external validity is more closely related to the realism created within the experiment than to the external trappings of similarity to real-world settings, which is referred to as mundane realism. As long as the experimental situation engages the subject in an authentic way, experimental realism has been constructed; under these circumstances, mundane realism may be nice but is hardly required to establish causality. Moreover, even if the experiment closely approximates real-world conditions, if its subjects fail to engage in an experimentally realistic way, subsequent findings are useless.

4. *Experimenter bias*. Experimenter bias, including expectancy effects and demand characteristics, can limit the relevance, generalizability, or accuracy of certain experimental results.

EXPERIMENTAL ETHICS

As a result of concerns about the ethical treatment of human subjects, the U.S. Department of Health and Human Services imposes strict guidelines on all research involving human subjects. Institutional review boards at major research institutions

oversee the administration of these guidelines and require advance approval on all experiments with human subjects to ensure their ethical treatment. These boards have the power to reject proposals that they deem to inadequately protect human subjects from unnecessary pain and suffering. The job of these boards involves weighing the risks and benefits of each study for the appropriate balance between risk to the subject and benefit for science or society. The guidelines include four important components:

1. *Informed consent.* Informed consent requires that the experimenter provide every subject with a disclosure statement prior to the experiment, describing the experimental procedures, along with expected gains and risks for the subject. Subjects are told they can leave the experiment at any time without penalty and are given contact information to report any concerns about their experience in the experiment to the institutional review board.
2. *Risk/gain assessment.* Experimenters are required to take every reasonable precaution to reduce the potential risks to their human subjects.
3. *Deception.* Some critics argue that any deception is not consistent with informed consent (Korn 1987), and others argue that deception is unethical and damages the reputation of experimental scientists in the larger society and is thus unjustified in any circumstance (Baumrind 1985). However, in some instances, deception may be necessary to ensure that subjects cannot guess the hypotheses under investigation. Were subjects to be aware of these hypotheses, their behavior would likely shift as a result of that knowledge, and results would be biased. When deception is used, institutional review boards are particularly careful to ensure that it is necessary. In addition, they may require experimenters to minimize its use, deny the application altogether, or require careful monitoring and reporting on the experiment while it is running.
4. *Debriefing.* At the end of an experiment, experimenters should be careful to tell the subjects as much as possible about the experiment. If deception was involved, experimenters should explain what the deception was and why it was deemed necessary for the unbiased collection of the data. In particular, subjects should be reassured that all their data will be confidential and that the experimenter will obtain subjects' written permission before any of their information is shared publicly.

REVIEW OF EXPERIMENTS RELEVANT TO POLITICAL SCIENCE

As noted, experiments have not been as widely employed in political science as in either psychology or behavioral economics. Why have we, as a discipline, been so slow to adopt experimental methodology? Why have all other social sciences used experiments to their advantage, and yet political scientists, with few exceptions,

resist incorporating experimental methodology into their own work or trusting those who do? Where has our field gone methodologically, and why has this movement been away from experimentation?

There are at least four reasons why political science has not been as receptive to the use of experiments as other social sciences. First, methodology in political science has moved toward large-scale multiple regression work. There is nothing wrong with this; indeed, in behavioral economics, there is a robust subfield devoted to using experiments in concert with formal models and statistical analysis to generate, test, and develop original hypotheses. So although experimental and formal or statistical methods need not be contradictory, the topics studied using these methods tend to be orthogonal to each other. For example, whereas multiple regression tends to concentrate on large groups, experimental work often focuses on small numbers of individuals.

Second, an alternative movement in political science, which tends to eschew the large-scale regression work, focuses on cultural and social aspects of particular phenomena. Although constructivists and postmodern scholars pursue some of the same issues as social psychologists, such as status concerns and the evolution of norms, most cultural analysts remain disinclined to believe that experimental work can get at phenomena as complex and multidimensional as political institutions or cultural and social structures.

The third concern is more practical. Lack of experimental training in political science at the graduate level means that few students add experimental design to their arsenal of research tools as a matter of course. As a result, only especially motivated students will contact an experimenter in a psychology or economics department to learn the basics of experimental design and procedure. The best way to learn how to do experiments is to run experiments. When training and experience are difficult or unavailable, the concentration of experimentalists required to shift disciplinary culture and practice fails to emerge.

Last, many political scientists, unlike other social scientists, seem to believe that experimentalists expect experimental work to stand on its own, the way it does in physics, chemistry, or biology. Unlike in biology, where every aspect of a particular investigation can take place within a petri dish, most phenomena of interest to political scientists are complex and involve many different variables. Experiments can be used very effectively, as they have been in experimental economics, to provide a middle ground between theory and naturally occurring empirical data.

Reciprocation between formal models and experimental testing has advanced both theory and method in behavioral economics, and can serve a similar function in political science as well. In addition, experimentation can be used to triangulate in on various issues for which other methods have produced either inconsistent or contradictory results. Finally, experiments can be effective in breaking down complex processes into isolated pieces, which can then be examined and explored in detail and in interaction. This process might demonstrate, for instance, that small changes in seemingly meaningless variables produce huge shifts in outcome, or that huge shifts in certain ostensibly important factors effect little change in

outcome. These insights, gained from the experimental procedure, help with theory development as well as theory testing.

Political science may have significant historical, cultural, or practical reasons for its lack of affinity to the use of experimentation; nonetheless, the past need not predict the future in this arena. Experimentation can achieve the same kind of successful impact in political science that it has had in other fields such as psychology and economics. We can learn from other social sciences that experiments need not stand on their own, as in biology, in order to be effective and useful in refining theory, providing evidence, and testing causal claims. Rather, experiments can dovetail with other methods in order to produce a cumulation of knowledge and an advancement in both theory and method within political science.

How might we respond to the previous lack of experimental work in political science? As noted, one of the mistakes that many political scientists make is to assume that experiments in political science need to mimic those in biology, if not in substance, at least in process. Yet, this stringent standard is not required; we need only note that experiments can dovetail with other methods to produce useful knowledge in order to adequately justify their utility.

In order to illustrate this process in action, the remainder of this essay is devoted to an explication of some of the experimental literature of relevance to mainstream political science. Certain experimental literature in behavioral economics and social psychology is addressed as well, since much of this work is relevant to issues of concern to political science. This article does not have the space to cover accomplishments in all experimental areas of interest to political science. Moreover, it is only through a sequence of experiments exploring related topics in different ways and on different populations that cumulative knowledge and external validity emerge. Therefore, following a brief overview of the presence of experiments in political science, I concentrate on systematic programs of larger research within behavioral economics and political science that have produced results with great relevance to major issues of interest in political science.

Overview

A comprehensive overview of experiments published by established political scientists reveals a total of 105 articles between 1926 and 2000. Only about 57 of them were published in political science journals. Many more strong articles written by political scientists, often in collaboration with economists, on political topics have been published in either psychology or economics journals. I examined 48 articles that appeared in non-political science journals but were written by political scientists; six or seven individuals, either alone or in collaboration, wrote the majority of them. The list of 105 articles does not include those published in the now-defunct *Experimental Study of Politics*. This journal was founded in 1971 because many believed that their experimental work had been unfairly rejected from the established political science journals (McConahay 1973); *Experimental Study* was created expressly to redress this difficulty. However, most of the articles published

in this journal were not as experimentally sophisticated as those published in psychology journals (McConahay 1973), and the journal ceased publication in 1975. A casual perusal of contemporary journals reveals that the same double standard exists today, with the experimental sophistication of typical articles in psychology or economics journals outstripping that of many articles published in political science journals. In addition, my review excluded some very good experiments published in edited volumes but not in journals (Iyengar & McGuire 1993, Palfrey 1991, Kagel & Roth 1995).

This overview revealed an increase in the number of published experiments over time. A single article was published in the 1920s, 5 in the 1950s, 7 in the 1970s, 42 in the 1980s, and 45 in the 1990s. (Note that the increase between the 1970s and the 1980s may have been at least partly due to the demise of the aforementioned *Experimental Study of Economics* journal in 1975.) Also interesting is the concentration of experiments in a small number of journals: *Public Opinion Quarterly (POQ)* and *American Political Science Review (APSR)* account for the most articles, with 21 each. These journals, with very few peripheral exceptions, were the only place that experiments were published in political science until the 1980s. *Political Psychology* has published 18 experimental articles since the 1980s. *American Journal of Political Science (AJPS)* has published 13 experiments and *Journal of Politics* has published 7. These five journals together account for 80 of the 105 experimental articles published by established political scientists in this period. Interestingly, the trend appears to be shifting, such that in the 1990s only 2 experimental articles were published in *POQ* and only 5 in *APSR*. *AJPS* captured the majority of the experimental articles published in the 1990s. Moreover, the remaining articles published in the 1990s appeared mostly in one of three additional journals [*Political Behavior (PB)*; *Journal of Risk and Uncertainty*; and *Games and Economic Behavior*], only one of which (*PB*) is primarily a political science journal.

It is quite striking that so few journals account for the majority of published experimental work in political science. The odds of getting an experimental article published drastically improve if it is submitted to one of these five journals. Although it is possible that political scientists simply do not submit experimental work to other political science journals by chance, the larger number of articles by political scientists on political topics published in either psychology or economic journals (*Econometrica* and *American Economic Review* in particular) suggest otherwise. The relatively low prevalence of experimental work in political science journals suggests that only a few are genuinely receptive to this form of methodological investigation. Indeed, personal communication from the editor of one major journal indicated that he could “never” put any experiment into print because he believed that experiments held no relevance whatsoever for political science. Such entrenched biases are difficult to overcome.

In addition, the bias toward experiments focused primarily on American topics is reflected in the predominance of voting behavior as a primary concern of experimental work. Out of the 105 articles surveyed, 25 were related to voting

behavior. The second, third, and fourth most popular topics were bargaining (13), games (10), and international relations topics (10). 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). Given the high percentage of experimental articles focused on voting, the question arises whether more experimenters actually focus on voting, or whether experiments on voting simply find more receptive journals available, and thus are more likely to be published overall.

Several areas of experimental research are relevant to the concerns of mainstream political scientists. Rather than examine each experiment individually, I discuss the major relevant findings in two broad areas of experimental research. In both of these areas, practitioners have developed sequences of experimental work on different populations using different measures in order to clarify and contextualize the generalizability of their results. Such sequences allow investigators to produce more reliable results, as one set of findings spurs another test that might narrow, reinterpret, or extend previous conclusions. Steady and gradual progress can thus attack seemingly impenetrable work. This essay concentrates on behavioral economics, political economy, and psychological studies of individual choice.

Experiments in Behavioral Economics

Within the past 15 years in particular, behavioral economics has grown exponentially and transformed itself through the systematic use of experimentation. These experiments often focus on issues of concern to political scientists. However, behavioral economics, like psychology, embraces some core assumptions about human nature and experimentation that not all political scientists share: skepticism toward notions of perfect rationality; emphasis on experimental validation of modeling assumptions; integration of micro level data; and adoption of lessons about human cognitive processing from experimental social and cognitive psychology (Laibson 2000). Further, experiments deemed successful in economics share certain central characteristics: clear instructions; absolutely no deception; stylized, stripped-down settings; anonymity; cash incentives; experimental tests of formal models, which can include comparisons with rational models; and, ideally, subsequent validation of the findings with supportive field data (Laibson 2000).

Obviously, economic topics have preoccupied political scientists using other methodologies, such as game theory and rational choice forms of analysis. In behavioral economics, theoretical development has profited from the interchange between formal models and experimental tests of such models. Formal models present hypotheses, which are then tested under experimental conditions. The experimental findings are used to refine and develop hypotheses to produce new theoretical models and predictions, which can then be experimentally tested in turn. Political scientists might benefit from this process as well in order to both establish empirical validation of formal models and speed the cumulation of knowledge.

Although the venues and contexts for investigation may differ, a great potential for substantive overlap exists between behavioral economics, psychology, and political science, which might be exploited for greater cross-disciplinary and interdisciplinary work. Current areas of research in psychology and economics offer promising opportunities for collaboration with political scientists who share such interests: 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).

Behavioral economics has concentrated on six main substantive areas since its experimental work began in the 1930s. These are (Roth 1995) (a) Prisoner's Dilemma and public goods issues; (b) problems of coordination and cooperation; (c) dynamics of bargaining; (d) experimental markets; (e) auction behavior; and (f) individual choice. Some experimental work has also been done in the area of industrial organization, although the majority of work in this area has been conducted by social psychologists.

PRISONER'S DILEMMA AND PUBLIC GOODS The Prisoner's Dilemma game was developed by Dresher and Flood at the Rand Corporation in 1950 (Flood 1952); the story was added by Tucker (1950) later (Straffin 1980). The central conundrum was that although cooperative play was transparently more profitable in the long run, equilibrium choice favored defection, producing less benefit for each player. Thus, the game produced a challenging test for equilibrium predictions. It was ideal for experimental investigation, which could examine alternative predictions in a controlled setting. In this environment, scholars developed a preference for repeated-play games. Most of these experiments demonstrated that cooperation begins early but breaks down over time, such that in multiround repeated games, players learn to defect earlier and earlier (Selten & Stoecker 1986). In most of these experiments, subjects know that a better outcome of mutual cooperation exists, but fear of exploitation makes mutual defection the only stable strategy over time. Even more interesting are recent findings suggesting that at least some players hold values independent of the payoffs embedded in the game structure, such as fairness, altruism, or concern with reputation building (Andreoni & Miller 1993). These findings will no doubt prompt further experimental work to test the nature and limits of these seemingly noneconomic motivations. Political science readers will also remember the famous Axelrod testing of Prisoner's Dilemma strategies in large-scale computer tournaments (Axelrod 1984). In his simulation, a "tit for tat" strategy, in which the player begins with cooperation and makes the same move as the opponent did the previous round, emerged as the most effective strategy for maximizing payoffs.

Apparently the public goods problem was first presented by Swedish economist Knut Wicksell in the nineteenth century (Roth 1995). Most experimental work on the public goods and free rider problem has focused on the conditions under which it might be most problematic and how its impact can be reduced in those situations.

Early work using single examples of public goods problems indicated little free riding (Johansen 1977). Like the work in Prisoner's Dilemma, experimental work on public goods problems moved toward repeated plays, demonstrating that in successive rounds, voluntary contributions decline (Isaac et al. 1985). Future work will no doubt explore the conditions under which free riding produces the greatest problems and examine ways in which to ameliorate its effects.

COORDINATION Economists investigating coordination problems experimentally study how subjects playing games with multiple equilibria decide on one in particular. In other words, how do players achieve a "meeting of the minds" (Ochs 1995)? Coordination clearly is influenced by particular aspects of the situation or environment, some of which are not adequately captured by traditional economic models. In many cases, coordination problems prevent optimal decision making from a rational perspective. For example, a factor that should not matter from a traditional economic perspective may exert a tremendous influence, as when the mere presence of a dominated strategy affects the equilibrium chosen (Cooper et al. 1990).

In an experiment involving only problems of coordination, with no inherent conflict of interest, Van Huyck et al. (1990) found that in a repeated game where outcomes are made public after each round, behavior quickly converges around the least profitable equilibrium. Obviously, this outcome does not represent economic rationality in the traditional sense. Crawford (1991) and others have sought to explain these findings using learning models and other game theory models offered by evolutionary biologists. Such models suggest that stable equilibria occur when strategies are not subject to invasion and dominance by new strategies. Crawford does not argue that the coordination problem itself is evolutionary in nature; rather, learning within the game allows stable, albeit economically nonrational, equilibria to emerge and dominate over time.

BARGAINING Experiments on bargaining behavior also examine how individuals arrive at an agreeable compromise equilibrium, with the added constraint of conflict of interest between players. According to most game theoretic models, player preferences and opportunities should determine outcomes of bargaining behavior. But experimental evidence has shown that bargaining outcomes are affected by variables that are not incorporated into player preference or opportunity, including the player's subjective expectations about the behavior of his opponent, which appears to be learned through experience (Roth & Schoumaker 1983).

The main debate in the bargaining literature has been between those who concentrate on game theoretic predictions and those who are interested in more sociological or cultural considerations. First, important findings in this area include the significance and frequency of disagreements and costly delays during the bargaining games. Second, many experiments demonstrate clear "deadline effects," which support the notion that many agreements are reached very close to the final deadline for agreement (Roth et al. 1988). Both these sets of findings have spurred additional work to determine the impact of these factors on the nature of bargaining

(Roth 1995). This topic is discussed below in the political science section on cooperation.

EXPERIMENTAL MARKETS Experimental markets typically draw on the basic design established by Chamberlain (1948) to encourage subjects to behave according to the laws of supply and demand, often with the goal of comparing alternative types of markets. Many of the important experiments in this area have been conducted by Plott and colleagues in response to specific policy questions surrounding the potential market effects of instituting particular government regulations (Hong & Plott 1982, Grether & Plott 1984).

Other work on experimental markets examines how traders use and share information. For example, work on asset valuation (Forsythe et al. 1982) demonstrates that prices tend to converge to a perfect equilibrium after replication. In addition, information aggregation (Forsythe & Lundholm 1990) has been examined within the context of experimental markets. Security markets have been examined in this way as well (Plott & Sunder 1988). Forsythe et al. (1992) used an experimental market whose ultimate value was tied to a future election outcome to examine how markets aggregate information. They found that their experimental market prices did a reasonable job of predicting the outcome of the election. On the basis of these results, Forsythe et al. (1992) argue that market transactions reduce the impact of biases such as political opinions on subjects' pricing decisions.

Provocative work on cross-cultural behavior in bargaining and experimental market environments suggested that market outcomes converged to equilibrium and that there were no payoff differences between subjects in Jerusalem, Ljubljana, Pittsburgh, and Tokyo (Roth et al. 1991). However, differences that deviated from equilibrium predictions did occur everywhere in both the agreement and the frequency of disagreement. The experimental procedures employed made the experimenters confident that observed differences in bargaining behavior did not result from differences in language or currency; rather, the investigators tentatively attributed these discrepancies to cultural differences.

AUCTIONS Auction behavior is familiar to anyone who is a fan of, or knows an addict of, eBay. The most consistent and prominent result of the study of auction behavior is the "winner's curse" (Bazerman & Samuelson 1983, Thaler 1992), wherein the winner of an auction pays much more for the prize than either its real value or the price he intended to pay at the outset. In addition, the average bid is often well below the objective value of the auctioned object. Uncertainty about the true value of the object appears to exacerbate this effect. These findings, suggestive of persistent mistakes on the part of bidders, are clearly contradictory to what established economic theory would predict about equilibrium behavior in auctions (Milgrom & Weber 1982).

INDIVIDUAL CHOICE Experimental work on individual choice has produced important findings in behavioral economics. This area of work retains the greatest

overlap between social psychologists and experimental economists in theory, method, and practice. Research on individual choice encompasses at least three important topics: preference reversals, judgment under uncertainty, and decision making under risk.

The topic of preference reversals has proved particularly problematic for economists attempting to sustain the accuracy of rational models (Tversky et al. 1990). Early experimental work by psychologists Slovic & Lichtenstein (1968; Lichtenstein & Slovic 1971) noted the anomaly that people will put a higher price on a given lottery when asked to buy or sell it (bids), but when asked to participate in one, most individuals will choose the other lottery (choice). These results remain remarkably robust. Preference reversals seem to result from differences in the way people process information about probabilities and payoffs. Bids appear to be governed by payoffs, whereas choices tend to be driven by probabilities. Although these findings remain inconsistent with expected utility and other rational models of decision making, psychologists readily explain this result as an example of the anchoring and adjustment heuristic (Kahneman et al. 1982), whereby people initially latch onto a value, which can be arbitrary and irrelevant, and fail to adequately adjust that value to present circumstances in making subsequent judgments. The first anchor people fix on is the monetary payoff, and then they insufficiently adjust choices to shifts in probabilities. Preference reversals result from this inadequate adjustment from the initial monetary anchor.

The second main area of research within the individual choice literature is judgment under uncertainty (Kahneman et al. 1982). This experimental work examines how individuals judge the frequency or likelihood of certain outcomes. This work consistently and robustly demonstrates at least three important judgmental heuristics that appear to control people's assessments of frequency: anchoring and adjustment (described above), representativeness, and availability. The representativeness heuristic claims that individuals assess frequency based on the similarity between the judged object or event and the categories to which it might belong. The availability heuristic argues that people judge likelihood based on salience, i.e., the ease of retrieval or imagination of the example from memory. All three judgmental heuristics contradict central assumptions in most expected utility models, which expect dominance, invariance, and intransitivity to hold sway in judgments about probabilities.

Decision making under risk has been most closely examined by the same psychologists who conducted the seminal work on judgmental biases (Kahneman & Tversky 2000). In attempting to develop a descriptively accurate model of choice as an alternative to expected utility models, Kahneman & Tversky (1979; Tversky & Kahneman 1992) delineated prospect theory. Prospect theory incorporates two successive phases: editing and evaluation. In editing, prospects or choices are framed for a decision maker. Robust experimental evidence indicates that trivial aspects of framing options can consistently exert profound impacts on the substance of choice. Specifically, seemingly trivial changes in the method, order, or form in which options are presented to a decision maker systematically affect the content of

choice. Evaluation itself encompasses two components as well: the value function and the weighting function. The value function has three central characteristics: (a) outcomes are judged in relative, not absolute, terms; (b) individuals tend to be risk-seeking in the domain of losses and risk-averse in the domain of gains; and (c) people tend to be loss-averse in general. The role of the weighting function is similar to, but distinct from, that of probability assessments in expected utility models. First, people have great difficulty incorporating extremes such as impossibility and certainty into their decision-making strategies. Second, individuals tend to overweight low probabilities while simultaneously underweighting moderate and high probabilities. All these empirical results surrounding the value and weighting functions contradict the predictions of standard expected utility models.

Experiments in Political Science

As in economics, some political scientists are beginning to use experimentation to test formal models in a controlled empirical setting. The three main areas of research in this area are (a) voting and elections, (b) committee and jury decision making, and (c) problems of coordination and cooperation (Palfrey 1991). This third area is similar in some theoretical ways to work on coordination in behavioral economics, but the domains of application differ. Palfrey (1991) argues that these topics have produced at least three important themes in the relationship between formal modeling and experimental research in political science. These include the importance of strategic behavior in studying complex political actions and actors; the critical significance of incomplete or asymmetric information, especially as related to issues of reputation, communication, and signaling; and finally, the importance of explicitly building dynamic models, which are aided by experimental methods and impact problems related to party identification, realignments, incumbency, and political business cycles.

EXPERIMENTAL VOTING AND ELECTIONS Since Downs's *An Economic Theory of Democracy*, many scholars have tried to examine the foundations of democratic elections with formal models. Increasingly, these models are being tested experimentally (Palfrey 1991). Plott (1991), for one, tested the spatial model to examine certain aspects of elections, including voter turnout.

Voter turnout lends itself nicely to experimental investigation. For example, Palfrey & Rosenthal (1985) argued that according to game theoretic analysis under assumptions of complete information, analysts should expect equilibria of high turnout, even when the costs of voting are high. Instead, they demonstrated experimentally that under conditions of uncertainty about the preferences and costs of others for voting, only voters with very low voting costs will vote in a large election. In other work on voter turnout, Green and colleagues have attempted to rehabilitate the use of field experiments begun by Gosnell (1926). In a study on the effects of canvassing, phone calls, and direct mail on voter turnout, Gerber

& Green (forthcoming¹) found that personal canvassing increased voter turnout, whereas phone calls appeared to have no impact. Direct mail appeared to have a slight impact on voter turnout. In addition, they found that asking voters whether they could be “counted on” to vote increased the impact of personal canvassing.

Other topics that have been investigated experimentally under the rubric of voting and elections include candidate competition (Plott 1991), retrospective voting (McKelvey et al. 1987), political competition (Boylan et al. 1991), and voter information costs.

Lau and Sears have used experiments to examine related topics. Their study of the evaluation of public figures (Lau et al. 1979) concluded that the so-called positivity bias often found in survey results is not an artifact of the measurement process alone but rests on some real bias in assessment. Related work on political preferences (Sears & Lau 1983) showed that self-interest may result from political and personal cues in surveys that trigger artifactual results. Finally, these authors have experimentally explored the nature of political beliefs (Lau et al. 1991).

Political party identification has also been examined experimentally (Cowden & McDermott 2000). We were intrigued by previous work, using different methodologies, that achieved somewhat contradictory results regarding the long-term stability of party identification. We designed an experiment that assessed student subjects' party identification, among other things, early in the semester. Later, after participating in one experiment that manipulated the extremity of real candidates in experimental elections, or another in which subjects role-played either the prosecutor or defender of Clinton in the impeachment hearing, subjects filled out a second, standard party identification measure. Our results indicated that party identification, even in a young population that should have had less time to develop strong associations, showed remarkable stability.

Media effects on candidate evaluation and voting have been another extremely productive research topic. Some of the best and most imaginative experimentation has been conducted in the area of media studies and political communication by Iyengar and colleagues. Their creative studies have demonstrated that television news influences how viewers weight problems and evaluate candidates (Iyengar et al. 1982); that television news frames individuals' explanation of events (Iyengar 1987); that negative advertising reduces voter turnout (Ansolabehere et al. 1994); and that candidates gain the most by advertising on issues over which they can claim “ownership” (Ansolabehere & Iyengar 1994). Iyengar continues to advance the methodology of experimentation itself as well, with recent studies that use new technology and field strategies to ameliorate some of the traditional criticisms of external validity problems (Iyengar 2000). These strategies include bringing the experiments into natural settings by creating living room environments in shopping malls and asking subjects to watch television in those settings, with experiments embedded in the programming. Further, Iyengar has begun to use the internet

¹Gerber A, Green D. *The Effects of Canvassing, Phone Calls and Direct Mail on Voter Turnout: A Field Experiment*. Unpublished manuscript.

to reach more diverse populations, which increases experimenter access to more representative samples.

Experimental studies of candidate evaluation by gender have produced some interesting findings as well. In an evaluation of campaign coverage of senatorial candidates, Kahn (1992) found that the press presented male and female candidates in systematically different ways. Such differences appeared to benefit male candidates, who were seen as more viable; this may disadvantage female candidates at the polls. Nevertheless, sex stereotypes sometimes benefit women because they were judged more frequently than men to be compassionate and honest. Further work by Kahn (1994) examining both gubernatorial and senatorial candidates found that voter perceptions were affected by both news coverage and sex stereotypes. Interestingly, these factors appear to affect incumbents differently from challengers, and gubernatorial candidates differently from senatorial candidates. In particular, gender differences in press coverage were more pronounced in the senate race and for incumbents. This pattern appears to hurt female senatorial candidates. On the other hand, sex stereotypes produce more positive evaluations of women and appear to benefit gubernatorial candidates the most. Note that Kahn's further experimental testing of her earlier findings allowed her to further refine and conditionalize her results. The findings of Huddy & Terkildsen (1993) on gender stereotyping in the perception of candidates are consistent with Kahn's. They too find that female candidates are seen in a positive light on traits such as compassion, whereas men are perceived to be more competent on military issues. Huddy & Terkildsen suggest that a gender trait approach best explains the differences they find.

COMMITTEE AND JURY DECISION MAKING A second arena of systematic research in political science considers committee and jury decision making. Experiments on committee decision making are typically modeled on legislatures in which results emerge from a combination of bargaining and voting. Much research thus focuses on how the bargaining process and the voting rules affect the outcome of committee decision making, especially under different decision rules. Various scholars have examined committee decision making under majority rule (Fiorina & Plott 1978, McKelvey & Ordeshook 1979), plurality (Neimi & Frank 1985), approval voting (Neimi 1984), noncooperative games (Felsenthal et al. 1988), competitive solutions (McKelvey & Ordeshook 1983), and universalism (Miller & Oppenheimer 1982). In particular, agenda setting (Levine & Plott 1977, Wilson 1986) and time constraints (Wilson 1986) offer perfect topics for experimental investigation based on strategic models. Guarnaschelli et al. (2000), among others, have recently used experimental work in the investigation of jury decision-making analysis as well.

Work on committee decision making often uses the ultimatum game as an experimental tool. Typically, two players must divide a sum of money, such as \$10. The procedure requires one player to offer an amount to the second, who can then accept or reject it. If the second player rejects it, no one gets the money, whereas if the player accepts it, both players split the money in the percentage agreed. The

theoretical question investigated is whether something about a subject's partner will affect either person's willingness to bargain.

Much work on committee decision making grew out of observations about the problems inherent in conventional game theory's treatment of these issues. The results of many experimental bargaining games seemed askew (Ostrom 1998) and players often exhibited consistent behavioral play (Camerer 1997), neither of which should be true according to traditional models. In particular, self-interest does not always work or dominate in these ultimatum games. This was also found in so-called dictator games, where the first player can solely dictate the division of goods. Specifically, unequal splits tend to be rejected in favor of "fair" splits. This outcome should not happen under subgame perfect backward induction equilibria, which would be predicted by expected utility models.

The promise of behavioral game theory rests on its ability to explore various aspects of this conundrum experimentally. In particular, experimental work can build on previous anomalous findings from ultimatum games to examine altruism, inequality aversion, and so-called mind reading (Wilson 2001). Altruism refers to a consistent desire to help others, even when it might hurt oneself. Numerous models of altruism (Forsythe et al. 1994, Eckel & Grossman 1996) typically assume that altruism is an embedded character trait within a given individual. Altruism succeeds because it gives people a positive feeling about themselves as a result of their actions. From an evolutionary standpoint, altruism may exist within communities because it advances the ability of the society to survive and prosper even when key individuals, such as mothers after childbirth, are too overwhelmed to perform their normal tasks successfully. Altruism in this sense may amount to little more than reciprocal selfishness.

Inequality aversion refers to many individuals' empirical preference for equal over unequal distributions of goods, even when extreme self-interest is possible, as in the dictator game. Examinations of this phenomenon explore how the same individual might act differently in different situations. What constraints control the extent to which an individual might cooperate in one circumstance but not in another? Early indications suggest that meaningful comparisons are important (Bolton & Ockenfels 2000) and that at least some inequality aversion derives from concerns surrounding relative status (Fehr & Schmidt 1999).

Mind reading refers to discerning the intentions of others (Rabin 1993; Levine 1998; Falk & Fischbacher, unpublished manuscript²). Unlike altruism, mind reading allows for the emergence of both positive and negative emotions (Frank 1988, Smith 1998). Once another person's intentions have been determined through an empirical process, people will tend to treat a nice person nicely and a mean person as they deserve.

Wilson (2001) has conducted a series of experiments investigating these phenomena. He demonstrates the ideal experimental procedures by learning from the

²Falk A, Fischbacher U. 1998. Kindness is the parent of kindness: modeling reciprocity. Unpublished manuscript.

experiences of previous studies and designing future studies to address past anomalies or to ameliorate procedural difficulties. He finds that beliefs about others are important and can change over time. These beliefs appear to be contingent on cues that individuals receive over time about others. In this way, interaction develops lasting reputations and labels. Wilson's work suggests that theoretical models of individual choice might be impaired by their failure to incorporate such seemingly nonrational factors as altruism, inequality aversion, and mind reading.

COORDINATION AND COOPERATION Work on coordination and cooperation in political science resembles similar work in behavioral economics, discussed above. However, applications differ, and work on cooperation in political science can easily be applied to problems in security as well as political economy. Topics include alliances, arms races, trade wars, and sanctions. For example, Geva and Skorick have used experimentation to test their cognitive calculus model of decision making in foreign policy (Geva & Skorick 1999, Geva et al. 2000, Geva & Skorick 2000). These authors use experimentation to test the predictions of their model against actual behavior in a laboratory setting.

Work on coordination and cooperation remains closely tied to work in social psychology and behavioral economics. Typically, scholars investigate this topic using noncooperative game theory (Palfrey 1991). Experimentalists seek to provide data related to certain models and push those models further by presenting evidence that might either refute or extend the current theoretical claims. Specific results indicate that communication increases group cooperation. Ostrom and colleagues (e.g., Ostrom & Walker 1991) have demonstrated that face-to-face communication, particularly in repeated-play settings involving common pool resources, exerts a powerful impact on propensity for cooperation.

Palfrey and colleagues have undertaken a systematic program of experimental research on topics related to coordination and cooperation. In one experiment, discounted repeated play proved more effective in generating cooperation than a single shot trial in a public goods game with incomplete information; however, results depended on the ability to monitor others and on the specific environmental conditions (Palfrey & Rosenthal 1994). Palfrey and colleagues have concentrated on the centipede game, in which two players alternately have a chance to take a larger portion of a continually escalating amount of money (McKelvey & Palfrey 1992, Fey et al. 1996). Once one person takes the money, the game ends. According to game theory predictions under assumptions of complete information, the first player should take the larger pile in the first round of play. However, this does not happen in reality. Rather, subjects operating under conditions of uncertainty and incomplete information about the payoff appear willing to consider the small possibility that they are playing against an altruistic opponent. Although the probability increases over time that a player will take the pile of money, the game typically continues into subsequent rounds. Palfrey has also investigated choice in other games (McKelvey & Palfrey 1995). This work shows great richness in its ability to combine formal modeling with experimental testing of such models.

The combination of methods allows greater confidence in results that point in the same direction.

Experimental work by Miller and colleagues has explored a variety of topics, including committees (Miller & Oppenheimer 1982). In work on games, Eavey & Miller (1984a) demonstrate that when universalist options, which offer “something for everyone,” exist in legislatures, concerns about fairness go beyond what expected value expectations would predict. Further, Miller & Oppenheimer (1982) find that competitive coalitions with a minimum winning coalition occur only when universal options are unavailable. In work on bargaining, Eavey & Miller (1984b) show that a bureaucratic monopoly on agenda setting allows bargaining with a voting body without necessarily imposing the agenda setter’s preferences on all. They conclude that bureaucratic agenda control in legislative bodies supports a bargaining model over an imposition one. Although some of this work (Miller & Oppenheimer 1982, Palfrey & Rosenthal 1994) points out the discrepancies between rational choice theory and the behavior of individuals in the real world, experiments are used not only to test and critique existing formal models but also to discover anomalies and challenges that are then incorporated into the next generation of model development.

Bolton (1991) has used experimentation to investigate how actual bargaining behavior differs from game theoretic predictions. Bolton & Zwick (1995) demonstrate that the opportunity to punish an opponent who treats you unfairly presents a more accurate explanation for deviations from perfect equilibrium solutions than the existence of anonymity for the subject. Note that although experimental findings may be at odds with some predictions of formal theory, the overall relationship between game theoretic modeling and experimentation in these exercises is collaborative; experiments empirically test formal models and suggest discrepancies as well as validations, and then formal modelers can attempt to incorporate these empirical demonstrations into later, more sophisticated models.

In our work on topics related to international relations, we investigate the impact of factors such as sex, uncertainty, and framing effects on arms races and aggression. In one experiment involving three rounds of a simulated crisis (McDermott & Cowden, forthcoming), we find that although uncertainty exerts no systematic effect on weapons procurement or likelihood of war, men are significantly more likely to purchase weapons and engage in aggressive action than women. In another experiment involving a simulated crisis game (McDermott et al. 2002), we examine the impact of framing in terms of striving for superiority or parity with the opponent, two kinds of uncertainty, and the tone of messages on weapons procurement. We find that embracing the frame of striving for superiority does indeed increase weapons procurement on the part of subjects. The tone of the message exerts a tremendous impact as well; recipients of hostile messages are much more likely to procure weapons than recipients of friendly messages. As in our other work, uncertainty appears to have no effect on weapons procurement. Finally, in more recent work, as yet unanalyzed, we manipulated the incentive to go to war to further examine the impact of sex differences on

levels of aggression. We plan to expand this paradigm to include other populations, including military officers, to further explore the impact of factors such as hormones, including testosterone, and nonverbal gestures on tendencies toward aggression.

CONCLUSIONS

Experimentation is one of many methods that can be used to examine political phenomena. Experiments have a long and distinguished history of effective usage in other disciplines, including hard sciences such as physics and biology, medicine, and social sciences such as psychology and economics. Unfortunately, experiments have been slower to acquire a dedicated following of practitioners in political science, mostly because of concerns about external validity. In many cases, this concern merely reflects a misunderstanding of the replication requirements necessary to establish external validity. But this concern may also indicate a failure to understand the difference between experimental realism, which is essential and requires the subject to be actively engaged in the process under investigation, and mundane realism, which refers to inessential trappings of the experimental situation that increase only the appearance, not the reality, of external validity (McDermott, under review³).

The primary advantage of experiments is that they offer unparalleled control over the variables of interest. This is because the experimental method permits the systematic manipulation of variables in a controlled environment with randomly assigned subjects. Experiments thus offer the highest degree of internal validity; experimenters can be pretty confident that outcomes differ on the basis of the variables manipulated systematically within the experimental conditions. This enables experimenters to make causal arguments about which factors cause certain outcomes, or contribute to them, and which do not.

Another advantage of experiments results from the scientific rigor built into the process. Experimenters remain aware of, and retain control over, the independent variables of interest. Experimenters carefully record results as dependent variables. Later statistical analysis allows the detailed testing of the relationships between these variables and any interactions among them. With this process, results that might not have been obvious to less systematic or large-scale analysis become prominent. Experiments allow causal inference, precise measurement and control, and clarity of detail.

Unfortunately, many political scientists assume that experimental results in political science need to be able to stand on their own, as in biology, and that if they cannot, they are useless. Nothing could be further from the truth. Experimentation can readily dovetail with other methodologies to produce systemic bodies of knowledge. As demonstrated by much of the work in behavioral economics and

³McDermott R. Experimental methodology in political science. Submitted.

some of the work in political science, the intersection of formal modeling and experimental testing is highly productive. Experiments can be, and have been, effectively used to test formal models, demonstrate unpredicted anomalies in outcomes that then provoke more sophisticated models, and suggest extensions and limitations of existing models under particular conditions.

In addition, experiments provide effective methodological help in examining areas in which other methodologies have produced inconsistent or contradictory findings, as was the case in our work on party identification. Experiments also offer clear advantages over other methods in particular areas of investigation, such as the validation of theories developed by formal modeling, or in further theory testing and refinement. Experiments offer useful insights in work that investigates the underlying process of a particular phenomenon as opposed to its outcome. Finally, invoking multiple methods, including experimentation, in investigating a phenomenon allows greater confidence in consensual results. In this way, experiments can help in triangulating in on research questions. Indeed, experimentation can serve a useful purpose, as it has in behavioral economics, to advance knowledge in political science more quickly and systematically and to cumulate such knowledge through the process of building on previous experimental work.

ACKNOWLEDGMENTS

I would like to thank Jonathan Cowden, Margaret Levi, one anonymous reviewer, the participants in the CBRSS Experimental Methods Conference at Harvard, and especially Sidney Tarrow for enormous help in writing this review.

The *Annual Review of Political Science* is online at <http://polisci.annualreviews.org>

LITERATURE CITED

- Andreoni J, Miller J. 1993. Rational cooperation in the finitely repeated Prisoner's Dilemma: experimental evidence. *Econ. J.* 103:570–85
- Ansolahehere S, Iyenger S. 1994. Riding the wave and claiming ownership over issues: the joint effects of advertising and news coverage in campaigns. *Public Opin. Q.* 58:335–57
- Ansolahehere S, Iyengar S, Simon A, Valentino N. 1994. Does attack advertising demobilize the electorate? *Am. Polit. Sci. Rev.* 88:829–38
- Aronson E, Carlsmith M. 1968. Experimentation in social psychology. In *The Handbook of Social Psychology*, ed. G Lindzey, E Aronson, Vol. 2. Reading, MA: Addison-Wesley. Rev. ed.
- Axelrod R. 1984. *The Evolution of Cooperation*. New York: Basic Books
- Baumrind D. 1985. Research using intentional deception: ethical issues revisited. *Am. Psychol.* 40:165–74
- Bazerman M, Samuelson W. 1983. I won the auction but don't want the prize. *J. Confl. Resolut.* 27:618–34
- Bolton G. 1991. A comparative model of bargaining: theory and evidence. *Am. Econ. Rev.* 81:1096–36
- Bolton G, Ockenfels A. 2000. Measuring motivations for the reciprocal responses observed

- in a simple dilemma game. *Am. Econ. Rev.* 90:166–93
- Bolton G, Zwick R. 1995. Anonymity versus punishment in ultimatum bargaining. *Games Econ. Behav.* 10:95–121
- Boylan R, Ledyard J, Lupia A, McKelvey R, Ordeshook P. 1991. Political competition in a model of economic growth: an experimental study. In *Laboratory Research in Political Economy*, ed. T Palfrey, pp. 33–68. Ann Arbor: Univ. Mich. Press
- Burke J, Greenstein F. 1989. *How Presidents Test Reality: Decisions on Vietnam, 1954 and 1965*. New York: Russell Sage Fdn.
- Camerer C. 1997. Progress in behavioral game theory. *J. Econ. Persp.* 11:167–88
- Campbell DT. 1968. Quasi-experimental design. In *International Encyclopedia of the Social Sciences*, ed. DL Sills, Vol. 5. New York: Macmillan
- Campbell DT, Ross HL. 1968. The Connecticut crackdown on speeding: time-series data in quasi-experimental analysis. *Law Soc. Rev.* 3:33–53
- Campbell DT, Stanley JC. 1963. Experimental and quasi-experimental designs for research on teaching. In *Handbook of Research on Teaching*, ed. NL Gage. Chicago: Rand McNally
- Chamberlain R. 1948. An experimental imperfect market. *J. Polit. Econ.* 56(2):95–108
- Cooper R, DeJong D, Forsythe R, Ross T. 1990. Selection criteria in coordination games: some experimental results. *Am. Econ. Rev.* 80:218–33
- Cowden J, McDermott R. 2000. Short term forces and partisanship. *Polit. Behav.* 22: 197–222
- Crawford V. 1991. An “evolutionary” interpretation of Van Huyck, Battalio and Beil’s experimental results on coordination. *Games Econ. Behav.* 3:25–59
- Eavey C, Miller G. 1984a. Fairness in majority rule games with a core. *Am. J. Polit. Sci.* 28: 570–86
- Eavey C, Miller G. 1984b. Bureaucratic agenda control: imposition or bargaining? *Am. Polit. Sci. Rev.* 78:719–33
- Eckel C, Grossman P. 1996. Altruism in anonymous dictator games. *Games Econ. Behav.* 16:181–91
- Etheredge L. 1978. *A World of Men: The Private Sources of American Foreign Policy*. Cambridge, MA: MIT Press
- Fehr E, Schmidt K. 1999. A theory of fairness, competition and cooperation. *Q. J. Econ.* 114:817–68
- Felsenthal D, Rapoport A, Maoz A. 1988. Tacit cooperation in three alternative noncooperative voting games: a new model of sophisticated behavior under the plurality procedure. *Elect. Stud.* 7:143–61
- Fey M, McKelvey R, Palfrey T. 1996. An experimental study of a constant-sum centipede game. *Int. J. Game Theory* 25:269–87
- Fiorina M, Plott C. 1978. Committee decisions under majority rule: an experimental study. *Am. Polit. Sci. Rev.* 72:575–98
- Flood M. 1952. *Some experimental games*. Res. Memo. RM-789, RAND Corp., June
- Forsythe R, Horowitz J, Savin N, Sefton M. 1994. Fairness in simple bargaining games. *Games Econ. Behav.* 6:347–69
- Forsythe R, Lundholm R. 1990. Information aggregation in an experimental market. *Econometrica* 58:309–48
- Forsythe R, Nelson F, Neumann G, Wright J. 1992. Anatomy of an experimental political stock market. *Am. Econ. Rev.* 82:1142–61
- Forsythe R, Palfrey T, Plott C. 1982. Asset valuation in an experimental market. *Econometrica* 50:537–67
- Frank R. 1988. *Passions Within Reason: The Strategic Role of Emotions*. New York: Norton
- Geva N, Mayhar J, Skorick JM. 2000. The cognitive calculus of foreign policy decision making: an experimental assessment. *J. Confl. Resolut.* 44:447–71
- Geva N, Skorick JM. 1999. Information inconsistency and the cognitive algebra of foreign policy decision making. *Int. Interact.* 25:333–62
- Geva N, Skorick JM. 2000. *Process and outcome consequences of simultaneous foreign policy decisions*. Presented at Annu. Meet.

- Int. Soc. Polit. Psychol., July 4–8, Seattle, WA
- Gosnell H. 1926. An experiment in the stimulation of voting. *Am. Polit. Sci. Rev.* 20:869–74
- Grether D, Plott C. 1984. The effects of market practices in oligopolistic markets: an experimental examination of the ethyl case. *Econ. Inq.* 22:479–507
- Guarnaschelli S, McKelvey R, Palfrey T. 2000. *An experimental study of jury decision rules*. Presented at Exp. Methods Conf., May 11–12, Cambridge, MA
- Hong J, Plott C. 1982. Rate filing policies for inland water transportation: an experimental approach. *Bell J. Econ.* 18:187–97
- Huddy L, Terkildsen N. 1993. Gender stereotypes and the perception of male and female candidates. *Am. J. Polit. Sci.* 37:119–47
- Isaac R, McCue K, Plott C. 1985. Public goods provisions in an experimental environment. *J. Public Econ.* 26:51–74
- Iyengar S. 1987. Television news and citizens' explanations of national affairs. *Am. Polit. Sci. Rev.* 81:815–32
- Iyengar S. 2000. *Experimental designs for political communication research: from shopping malls to the internet*. Presented at Exp. Methods Conf., May 11–12, Cambridge, MA
- Iyengar S, McGuire W, eds. 1993. *Explorations in Political Psychology*. Durham, NC: Duke Univ. Press
- Iyengar S, Peters MD, Kinder D. 1982. Experimental demonstrations of the “not-so-minimal” consequences of television news programs. *Am. Polit. Sci. Rev.* 76:848–58
- Johansen L. 1977. The theory of public goods: misplaced emphasis? *J. Public Econ.* 7:147–52
- Kagel J, Roth A, eds. 1995. *The Handbook of Experimental Economics*. Princeton, NJ: Princeton Univ. Press
- Kahn K. 1992. Does being male help? An investigation of the effects of candidate gender and campaign coverage on evaluations of U.S. senatorial candidates. *J. Polit.* 54:497–517
- Kahn K. 1994. Does gender make a difference? An experimental examination of sex stereotypes and press patterns in statewide campaigns. *Am. J. Polit. Sci.* 38:162–95
- Kahneman D, Slovic P, Tversky A, eds. 1982. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge, UK: Cambridge Univ. Press
- Kahneman D, Tversky A. 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 47:263–91
- Kahneman D, Tversky A. 2000. *Choices, Values and Frames*. New York: Cambridge Univ. Press/Russell Sage Fdn.
- Korn J. 1987. Judgments of acceptability of deception in psychological research. *J. Gen. Psychol.* 114:205–16
- Kuklinski J, Sniderman P, Knight K, Piazza T, Tetlock P, et al. 1997. Racial prejudice and attitudes toward affirmative action. *Am. J. Polit. Sci.* 41:402–19
- Laibson D. 2000. Untitled. Presented at Exp. Methods Conf., May 11–12, Cambridge, MA
- Lau R, Sears D, Centers R. 1979. The “positivity bias” in evaluations of public figures: evidence against the instrument artifacts. *Public Opin. Q.* 43:347–58
- Lau R, Smith R, Fiske S. 1991. Political beliefs, policy interpretations, and political persuasion. *J. Polit.* 53:644–75
- Levine D. 1998. Modeling altruism and spitefulness in experiments. *Rev. Econ. Dyn.* 1:593–622
- Levine M, Plott C. 1977. Agenda influence and its implications. *Va. Law Rev.* 63:561–604
- Lichtenstein S, Slovic P. 1971. Reversal of preferences between bids and choices in gambling decisions. *J. Exp. Psychol.* 89:46–55
- McConahay J. 1973. Experimental research. In *Handbook of Political Psychology*, ed. JN Knutson, pp. 356–82. San Francisco: Jossey-Bass. 542 pp.
- McDermott R, Cowden J. 2001. The effects of sex and uncertainty in a crisis simulation game. *Int. Interact.* 27:353–80
- McDermott R, Cowden J, Koopman C. 2002. The effects of framing, uncertainty and hostile communications in a crisis simulation game. *Polit. Psychol.* In press

- McKelvey R, Ordeshook P. 1979. An experimental test of several theories of committee decision making under majority rule. In *Applied Game Theory*, ed. S Brams, A Schotter, G Schwodiauer. Wurzburg, Ger.: Physica Verlag
- McKelvey R, Ordeshook P. 1983. Some experimental results that fail to support the competitive solution. *Public Choice* 40:281–91
- McKelvey R, Ordeshook P, Collier K, Williams K. 1987. Retrospective voting: an experimental study. *Public Choice* 53:101–30
- McKelvey R, Palfrey T. 1992. An experimental study of the centipede game. *Econometrica* 60:803–36
- McKelvey R, Palfrey T. 1995. Quantal response equilibria for normal form games. *Games Econ. Behav.* 10:6–38
- Milgrom P, Weber R. 1982. Predation, reputation, and entry deterrence. *J. Econ. Theory* 27:280–312
- Miller G, Oppenheimer J. 1982. Universalism in experimental committees. *Am. Polit. Sci. Rev.* 76:561–74
- Neimi R. 1984. The problem of strategic behavior under approval voting. *Am. Polit. Sci. Rev.* 78:952–58
- Neimi R, Frank A. 1985. Sophisticated voting under the plurality procedure: a test of a new definition. *Theory Decision* 19:151–62
- Ochs J. 1995. Public goods. In Kagel & Roth 1995, pp. 195–253
- Ostrom E. 1998. A behavioral approach to the rational choice theory of collective action. *Am. Polit. Sci. Rev.* 92:1–22
- Ostrom E, Walker J. 1991. Communication in a commons: cooperation without external enforcement. In Palfrey 1991, pp. 287–322
- Palfrey T, ed. 1991. *Laboratory Research in Political Economy*. Ann Arbor: Univ. Mich. Press. 323 pp.
- Palfrey T, Rosenthal H. 1985. Voter participation and strategic uncertainty. *Am. Polit. Sci. Rev.* 79:62–78
- Palfrey T, Rosenthal H. 1994. Repeated play, cooperation and coordination: an experimental study. *Rev. Econ. Stud.* 61:545–65
- Plott C. 1991. A comparative analysis of direct democracy, two-candidate elections, and three-candidate elections in an experimental environment. In *Laboratory Research in Political Economy*, ed. T Palfrey, pp. 11–32. Ann Arbor: Univ. Mich. Press
- Plott C, Sunder S. 1988. Rational expectations and the aggregation of diverse information in laboratory security markets. *Econometrica* 56:1085–118
- Rabin M. 1993. Incorporating fairness into game theory and economics. *Am. Econ. Rev.* 83:1281–302
- Roethlisberger FJ, Dickson WJ. 1939. *Management and the Worker*. Cambridge, MA: Harvard Univ. Press
- Rosenberg MJ. 1965. When dissonance fails: on eliminating evaluation apprehension from attitude measurement. *J. Pers. Soc. Psychol.* 1:28–42
- Rosenthal R. 1966. *Experimental Effects in Behavioral Research*. New York: Appleton-Century-Crofts
- Roth A. 1988. Laboratory experimentation in economics: a methodological overview. *Econ. J.* 393:974–1031
- Roth A. 1995. Introduction. In Kagel & Roth 1995, pp. 3–109
- Roth A, Murnighan K, Schoumaker F. 1988. The deadline effect in bargaining. *Am. Econ. Rev.* 78:806–23
- Roth A, Prasnikar V, Okuno-Fujiwara M, Zamir S. 1991. Bargaining and market behavior in Jerusalem, Pittsburgh, and Tokyo: an experimental study. *Am. Econ. Rev.* 81:1068–95
- Roth A, Schoumaker F. 1983. Expectations and reputations in bargaining: an experimental study. *Am. Econ. Rev.* 73:362–72
- Sears D. 1986. College sophomores in the laboratory: influences of a narrow data base on social psychology's view of human nature. *J. Pers. Soc. Psychol.* 86:515–30
- Sears D, Lau R. 1983. Inducing apparently self-interested political preferences. *Am. J. Polit. Sci.* 27:223–52
- Selten R, Stoecker R. 1986. End behavior in sequences of finite Prisoner's Dilemma

- supergames: a learning theory approach. *J. Econ. Behav. Org.* 7:47–70
- Slovic P, Lichtenstein S. 1968. Relative importance of probabilities and payoffs in risk taking. *J. Exp. Psychol. Monogr. Suppl.* 78:1–18
- Smith V. 1998. The two faces of Adam Smith. *South. Econ. J.* 65:1–19
- Sniderman P, Piazza T, Tetlock P, Kendrick A. 1991. The new racism. *Am J. Polit. Sci.* 35:423–47
- Straffin P. 1980. The Prisoner's Dilemma. *UMAP J.* 1:102–3
- Thaler R. 1992. *The Winner's Curse: Paradoxes and Anomalies of Economic Life*. Princeton, NJ: Princeton Univ. Press
- Tucker A. 1950. A two-person dilemma. Mimeo, Stanford Univ.; republished 1980. On jargon: the Prisoner's Dilemma. *UMAP J.* 1:1
- Tversky A, Kahneman D. 1992. Advances in prospect theory: cumulative representation of uncertainty. *J. Risk Uncertain.* 5:297–323
- Tversky A, Slovic P, Kahneman D. 1990. The causes of preference reversal. *Am. Econ. Rev.* 80:204–17
- Van Huyck J, Battalio R, Beil R. 1990. Tacit coordination games, strategic uncertainty, and coordination failure. *Am. Econ. Rev.* 80:234–48
- Walker T. 1976. Microanalytic approaches to political decision-making. *Am. Behav. Sci.* 20:93–110
- Wilson R. 1986. Forward and backward agenda procedures: committee experiments on structurally induced equilibrium. *J. Politics* 48:390–409
- Wilson R. 2001. Untitled. Presented at Exp. Methods Conf., May 11–12, Cambridge, MA
- Zimbardo P, Gerrig R. 1996. *Psychology and Life*. New York: HarperCollins. 14th ed.

Copyright of *Annual Review of Political Science* is the property of Annual Reviews Inc. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.