

# The Growth and Development of Experimental Research in Political Science

JAMES N. DRUCKMAN *Northwestern University*

DONALD P. GREEN *Yale University*

JAMES H. KUKLINSKI *University of Illinois at Urbana-Champaign*

ARTHUR LUPIA *University of Michigan*

**A**lthough political scientists have long expressed skepticism about the prospects for experimental science, an analysis of the first hundred volumes of the *American Political Science Review* reveals that randomized experiments have grown in impact and prominence. We document how thinking about experimentation has evolved over the century, and demonstrate the growing influence of laboratory, survey, and field experiments. A number of experiments have transformed how political scientists think about causal relationships in specific substantive areas. There are limits to the kinds of questions that experiments can address, but experiments have made important contributions in an array of political science subfields.

**I**n his 1909 APSA presidential address, A. Lawrence Lowell advised the fledgling discipline against following the model of the natural sciences. “We are limited by the impossibility of experiment. Politics is an observational, not an experimental science ...” (Lowell 1910, 7). The lopsided ratio of observational to experimental essays in the *Review*’s first hundred volumes arguably affirms Lowell’s assessment, but the next hundred volumes are likely to be quite different. The number and influence of experimental studies is growing rapidly, as political scientists discover ways of using experimental techniques to illuminate political phenomena.

The drift toward experimentation reflects the value that our discipline places on causal inference and empirically guided theoretical refinement. Experiments facilitate causal inference through the transparency and content of experimental procedures, most notably the random assignment of observations to control and treatment groups. They also guide theoretical development by providing a means for pinpointing the effects of institutional rules, preference configurations, and other contextual factors whose impact can be difficult to gauge using other forms of inference.

Although the range of laboratory, survey, and field experimentation has developed rapidly, any assessment of experimental political science must acknowledge its limitations. To date, the range of application

remains narrow, with most experiments pertaining to questions in the subfields of political psychology, electoral politics, and legislative politics. An important question is the extent to which experiments or experiment-inspired research designs can benefit other subfields.

We use a combination of qualitative and quantitative methods to characterize the growth, development, and limitations of experimentation in political science. We begin by presenting a content analysis that traces how the meaning of the term “experiment” has evolved over the past century. Only recently have political scientists come to construe the experiment as members of many other disciplines do, namely, as a scientific methodology with an unrivaled capacity to demonstrate cause and effect; and thus only recently have political scientists begun to use experimental designs in the laboratory, in surveys, and in field studies. The increased use of experiments in recent years, however, has been dramatic.

To understand the recent emergence of experiments, we invoke economist Alvin Roth’s three purposes of experiments: to search for facts, speak to theorists, or “whisper in the ears of princes.” Using this classification scheme, we suggest that the rise of experimental designs in political science reflects both an increased demand to test causal claims that existing methodologies could not fully address and an increased capacity of experiments to simulate relevant political contexts. We then describe in detail a sample of influential experiments that have contributed to ongoing substantive, theoretical, and policy debates and that illustrate our general claims about changing demand and supply. We next use a quantitative analysis of citation counts to assess the contribution of experimental articles to scholarly discourse. We find that experimental articles are in fact cited with greater frequency than other contemporaneous articles. Finally, we conclude with remarks about the limited substantive domains on which experimental articles have concentrated and the prospect for experimental reasoning to affect other parts of the discipline.

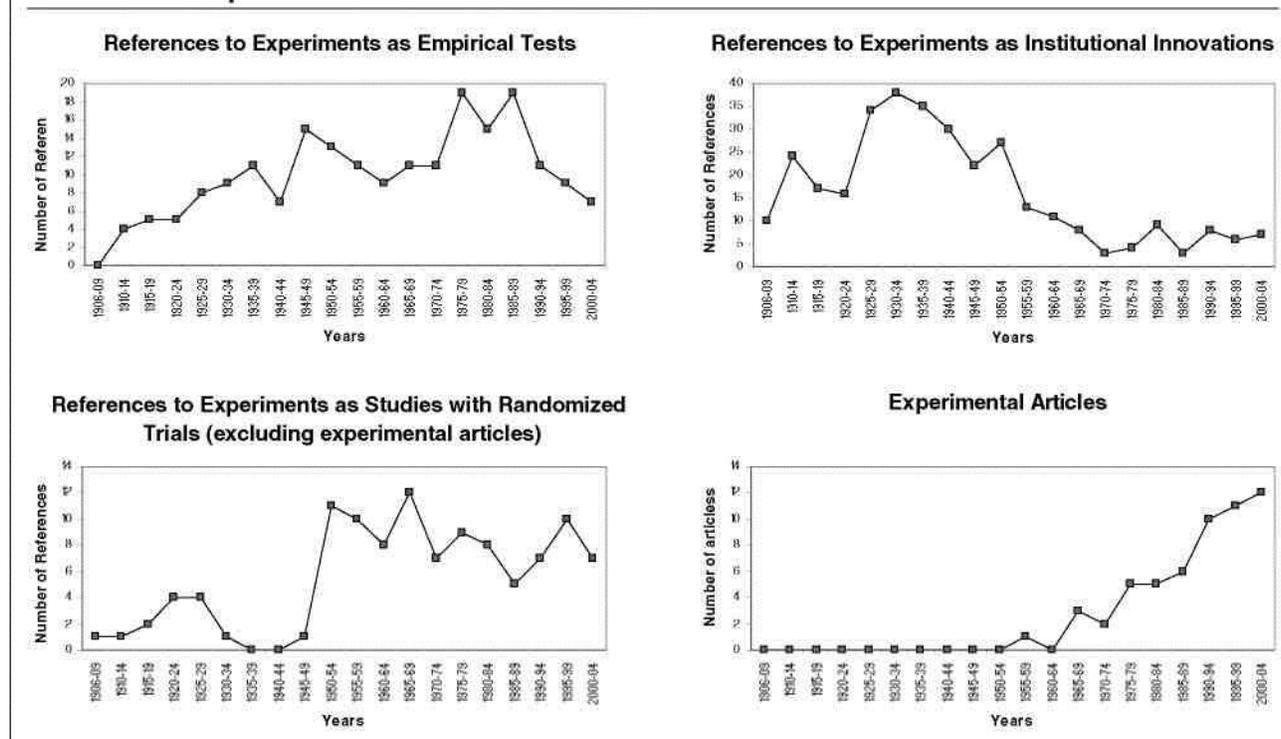
James N. Druckman is an Associate Professor, Department of Political Science, Northwestern University, Scott Hall, 601 University Place, Evanston, IL 60208 (druckman@northwestern.edu).

Donald P. Green is A. Whitney Griswold Professor, Yale University, New Haven, CT 06520.

James H. Kuklinski is Matthew T. McClure Professor, University of Illinois at Urbana-Champaign, Urbana, IL 61801.

Arthur Lupia is Hal R. Varian Collegiate Professor, University of Michigan, Ann Arbor, MI 48109.

We thank the editors, the reviewers, and John Bullock for helpful comments, and Toby Bolsen, Kjersten Nelson, Bas van Doorn, Tony Evans, Samantha Holland, Jason Kustka, Jackie Len, and John Wainwright for research assistance. We also thank Tobin Grant for providing a list of quantitative review articles. Additional information on the content analyses is available from the authors.

FIGURE 1. “Experimental” Trends in the *Review*

## THE CHANGING MEANING OF EXPERIMENTATION IN POLITICAL SCIENCE

To most contemporary political scientists, the term “experiment” connotes a particular scientific methodology. This has not always been the case. To assess how political scientists have used the term “experiment” over the past 100 years, we tracked every use of the word in the *Review*.<sup>1</sup> The analysis indicates that authors have used the term in three distinct ways, referring to (1) an *institutional innovation* such as a new constitution, electoral system, or policy process; (2) a *simulation or an empirical test* that involves neither an institution nor randomized trials; and (3) a *randomized trial* in which the researcher randomly assigns units of observation to control and treatment groups. Figure 1 reports the numbers of references to the three usages by five-year periods.

From the beginning of the *Review* through the 1950s, authors used the term “experiment” largely to describe changes to governmental institutions, reflecting the formal legalism that dominated scholarly debate. This usage ran counter to, and completely ignored, an emergent conception of experiments associated with the work of statistician R. A. Fisher (1935). By 1951, Harold Lasswell, in his *Review* essay (1951, 139),

<sup>1</sup> We used JSTOR to do the search. A 30% random sample of usages was coded by a second coder, and we found 95% agreement between coders (correcting for chance agreement). We exclude papers that present an actual randomized experiment from the word search results.

was expressing frustration with the continued use of “experiment” in the pre-Fisher sense:

In popular usage it is common to speak of a limited official introduction of a new practice as an “experiment,” as when it is said that the States have experimented with new institutions of government. Strictly, the term experiment is too useful to be applied in this way. We need a word with the scientific connotation of “controlled manipulation,” as when we modify a single factor, and hold everything else constant.

Lasswell’s expressed sentiment anticipated the shift in outlook that accompanied the rise of behavioralism in the 1950s. Figure 1 shows that references to experiments as institutional arrangements dropped dramatically around that time. Henceforth, scholars overwhelmingly used the term to refer to nonexperimental empirical tests or to true random-assignment experiments. It was not until the 1990s, however, that nonexperimental papers began to routinely refer to experiments as actual experiments rather than as other kinds of empirical tests. This trend reflects a convergence between political science and other disciplines that see experiments as a well-defined, accepted, and important methodology.

The last panel in Figure 1 displays the temporal pattern in the number of published *Review* articles using randomized experiments over the journal’s first hundred years. That number—57 in total—has grown sharply in recent years.<sup>2</sup> The first experiment appeared

<sup>2</sup> Our list of articles builds directly on McGraw and Hoekstra 1994; otherwise, we require that the article include the presentation of

in 1956. The second did not appear until a decade later. Although a steady stream of experimental articles appeared in the 1980s, more than half of the articles whose research fits the conventional definition of an experiment appeared after 1992.<sup>3</sup> *Research satisfying the modern definition of experiment is a relatively new phenomenon in political science.*

The changes in demand and supply that we document below explain these evolutionary dynamics. They become evident, however, only in terms of the purposes that experiments serve. As Roth (1995) has noted, researchers use experiments for multiple purposes.

## THE PURPOSES OF EXPERIMENTS

Roth (1995, 22) identifies three nonexclusive roles that experiments can play. One is what he describes as “searching for facts,” where the goal is to “isolate the cause of some observed regularity, by varying details of the way the experiments were conducted. Such experiments are part of the dialog that experimenters carry on with one another.” These types of experiments often complement observational research by arbitrating between conflicting results (i.e., inconsistent relationships between variables) documented in observational data. “Searching for facts” describes many experimental studies that attempt to estimate the magnitudes of causal parameters, such as the influence of racial attitudes on policy preferences (Gilens 1996) or the price-elasticity of demand for public and private goods (Green 1992). A second role entails “speaking to theorists,” where the goal is “to test the predictions [or the assumptions] of well articulated formal theories [or other types of theories] . . . Such experiments are intended to feed back into the theoretical literature—that is, they are part of a dialogue between experimenters and theorists.” The many political science experiments that assess the validity of claims made by formal modelers epitomize such correspondence (e.g., Fréchette, Kagel, and Lehrer 2003, Morton 1993, Ostrom, Walker, and Gardner 1992).<sup>4</sup> The third usage is “whispering in the ears of princes,” which facilitates “the dialogue between experimenters and policymakers . . . [The] experimental environment is designed to resemble closely, in certain respects, the naturally occurring environment that is the focus of interest for the policy purposes at hand.” Cover and Brumberg’s (1982) field experiment examining the effects of franked mail on constituent opinion is an ex-

primary data from a random assignment study with participants. Coders content analyzed every experiment in each article, with average intercoder agreement of near 90%. Our coding is based on our reading of what the authors explicitly report. We do not include “natural experiments” in which near-random assignment occurs by some natural process or discontinuity.

<sup>3</sup> The initial rise of experiments in the late 1960s and the early 1970s undoubtedly contributed to the publication of a short lived journal entitled *The Experimental Study of Politics*.

<sup>4</sup> However, the theories need not be formal; for example, Lodge and his colleagues have implemented a series of experiments to test psychological theories of information processing (e.g., Lodge, McGraw, and Stroh 1989; Lodge, Steenbergen, and Brau 1995).

ample of an experiment that whispers in the ears of congressional “princes.”

Although political scientists might share rationales for experimentation with other scientists, what distinguishes their efforts is their attention to focal aspects of politically relevant contexts. This distinction parallels the use of other modes of inference by political scientists. As Druckman and Lupia (2006, 109) argue, “[c]ontext, not methodology, is what unites our discipline . . . Political science is united by the desire to understand, explain, and predict important aspects of contexts where individual and collective actions are intimately and continuously bound.” The environment in which an experiment takes place is thus of particular importance to political scientists.

## CHANGES IN DEMAND AND SUPPLY

The preceding purposes, when put in terms of demand and supply, make clear why experimental research in political science has grown in recent years. The initial turn toward experimentation in the 1960s and 1970s coincided with the rise of behaviorism, some of which drew on psychology, where experiments are the central mode of inquiry. During the 1980s and beyond, the influence of psychology, and thus experiments, became even greater as a younger group of scholars borrowed heavily from social cognition to define the distinct field of political psychology (e.g., Druckman and Miller 2004).

During the same period, scholars began to assemble quantitative data sets and apply advanced statistical methods to them to uncover causal relationships. For many topics, these observational methods yielded important substantive advances. In other cases, limitations stemming from such issues as endogeneity, self-selection, and omitted variable bias resulted in ambiguous and debated correlations. Growing sensitivity to such matters led some political scientists to explore experiments as a method for identifying causal relationships. This same period also saw the increasing use of formal theory to derive *ceteris paribus* causal arguments. Empirical assessments of these arguments fueled greater demand for laboratory experimentation because the controlled environment of the laboratory enabled researchers to structure institutional settings and participant incentives in ways that correspond to a model’s structure.

The provision of experiments depends not only on demand but also on supply. Methodological and technological advances—such as induced value theory (where experimentalists use financial incentives to control subjects’ preferences) and computer-assisted telephone interviews (CATI)—have enabled researchers to use experiments in a wide array of politically relevant contexts by either experimentally creating settings or inserting experiments into natural-occurring environments. For example, the rise in experiments that started in the early 1990s coincided with the increasing availability of CATI technology to implement survey experiments (Sniderman and Grob 1996). Nearly two-thirds of the survey experiments published in the

*Review* occurred after 1992. Another contributing factor has been the increasing availability of experimentally relevant software, such as programs that facilitate the creation of strategic environments akin to game theoretic models and technology that enables researchers to edit various forms of media presentations.

Of equal, if not greater, importance than methodological and technological developments is the creativity and originality shown by researchers in making use of these tools (see Sniderman 1995 for discussion). For instance, the growth of survey experiments in the 1990s might not have occurred had it not been for the inventive stewardship of scholars such as Paul Sniderman and his colleagues (e.g., Sniderman, Brody, and Tetlock 1991) who initiated the multi-investigator experimental surveys (the predecessor to Timesharing Experiments in the Social Sciences).<sup>5</sup> Likewise, Charles Plott, Richard McKelvey, and Peter Ordeshook, among others, were at the forefront of using experiments to test formal theories, while Shanto Iyengar and Donald Kinder introduced the use of laboratory experiments to explore the impact of the media. These are but a few of many examples of imaginative scholars who envisioned novel experimental approaches to studying pressing political questions.

## A QUALITATIVE ANALYSIS OF SELECTED EXPERIMENTAL STUDIES

The following qualitative review shows how the general changes in demand and supply set the stage for a sample of particularly prominent experiments. It provides detailed examinations of how these changes spurred researchers to adopt experiments to address perennial research questions. We organize our review by offering examples of experiments whose primary goal fits one of Roth's categories (also see McDermott 2002; Morton and Williams 2006).

### Searching for Facts

Iyengar, Peters, and Kinder's (1982) study of the effects of television news represents one of the best-known studies in the "searching for facts" category. Prior to it, many non-experimental tests reported strong correlations between the issues on which the news media focused and the issues that citizens believed were important. The news seemed to set citizens' agendas. Yet the evidence for agenda setting was problematic because observational data could not sort out causal direction. For example, are correlations between the issues that the news media report and the issues that citizens identify as important evidence of the media leading public attention in certain directions, or do the correlations reflect the news media and the audience simultaneously responding to real-world events?

<sup>5</sup> Sniderman's own work on racial attitudes and political reasoning has strongly influenced the study of public opinion. See Sniderman, Hagendoorn, and Prior 2004 for a recent *Review* example of that work

Iyengar, Peters, and Kinder's (1982) implemented an ingenious set of laboratory experiments that randomly assigned participants to newscasts that differed only in terms of the issues covered. They used actual newscasts for the stimuli, invited nonstudent participants to watch and evaluate the news in a realistic group setting that resembled a living room, embedded the key outcome measures in a lengthy questionnaire, and measured responses up to 24 hours after exposure. The results showed that participants' opinions about issue importance substantially reflected the content of the specific broadcast to which they were exposed, indicating a clear causal relationship.

This set of experiments offered persuasive evidence of the media's potential impact and initiated a research paradigm that has generated a host of related experiments, several of which have appeared in the *Review*. Like Iyengar and colleagues (1982) subsequent experimenters have pinned down causal relationships about opinion formation in ways not possible with nonexperimental data (e.g., Ansolabehere et al. 1994; Lau and Redlawsk 1997; Mutz 2002; Nelson, Clawson, and Oxley 1997). The experiments also stimulated nonexperimental research exploring different types of media effects (e.g., Bartels 1993)—something that had rarely been done in the discipline prior to Iyengar et al.'s work.

A second example centers on the thesis that advanced industrial societies are shifting from materialist to post-materialist values. Presenting across-time aggregate survey responses to a battery of value questions from Euro-Barometer Surveys, Inglehart (1971, 1977, 1990) has argued that economic concerns are becoming less important in these countries. Critics claimed that Inglehart's findings were an artifact of prevailing strong economic conditions. This debate, carried on largely with aggregate-level time-series data, appeared to have reached a stalemate.

In paired survey experiments, Clarke and colleagues (1999) tested the performance of the Euro-Barometer values battery in Canada and Germany, two countries experiencing high unemployment and low inflation at the time of the study. Most crucially, they demonstrated that substituting an unemployment item for the standard inflation item dramatically changed the results: far fewer people could be categorized as postmaterialist. Much of the documented shift from materialism to postmaterialism, the authors conclude, is a measurement artifact.

Two aspects of this study warrant special comment. First, the Clarke et al. (1999) experiments offered far more compelling evidence about the validity of the Euro-Barometer items than any observational study had been able to do. Only an experimental design could convincingly demonstrate how substitution of one survey item for another dramatically changed respondents' value classifications. It took an experiment, in other words, to reveal this causal connection. Second, Clarke et al. deployed their experiment in cross-national contexts with high unemployment rates, showing that the effects of question wording were not specific to a particular setting; rather they reflected

a more general psychological phenomenon of broad importance to comparative survey analysis.

## Speaking To Theorists

A second reason for conducting experiments is to assess the empirical validity of a theoretical proposition. In such experiments, the focal question is “If we can construct an experimental setting in which we can vary whether certain assumptions of a theory are true or false, will we observe outcomes that are consistent with theoretical propositions?” To reach such an assessment, the goal is to make the experimental setting resemble, as closely as possible, the setting posited in the theory. Whether either setting corresponds to some reader’s conception of “the real world” might or might not be an interesting question, but it is not the point of this kind of experimental endeavor (Cook and Campbell 1979, 83).

A well-known example of this kind of experiment is Fiorina and Plott’s (1978) study of committee decision making, in which undergraduate subjects were randomly assigned to ideal points and were rewarded according to the closeness between their ideal points and the final decision ratified by committee vote. Across a wide array of experimental manipulations, Fiorina and Plott found that the predictions derived from formal models of utility maximizing agents outperform those derived from alternative sociological theories. Experimental studies in this genre—where experimenters can control the institutional setting and shape participant’s preferences—had an enormous influence on the subsequent literature on committees, which largely abandoned the sociological theories in favor of models that emphasize strategic behavior and institutional context.

Other experiments that speak to theorists are concerned with the correspondence between theory and reality. One such experiment, by Quattrone and Tversky (1988), is the most widely cited experimental article published in the *Review*. The authors examined the assumption that decision makers place equal and opposite value on gains and losses of equivalent size, but contended instead that subjects are more sensitive to losses than to gains. Using a between-subjects design, they presented undergraduates at Stanford and Berkeley with a series of hypothetical choices, some framed as gains and others as losses. In apparent violation of basic axioms of rationality, subjects appeared to experience “preference reversals”: they chose A over B when the consequences were presented in one way, chose B over A when they were presented another way, and made these choices despite the fact that the consequences of choosing A and B were materially equivalent.

For example, Quattrone and Tversky (1988) examine a rationality axiom that builds on the assumption that people’s preferences toward an economic program will not depend on whether that program is said to result in 90% employment or 10% unemployment. In fact, though, preferences depend on how the program is framed: people are more supportive when employment is emphasized. The critical element of the experiment

was that Quattrone and Tversky were able randomly to present some participants with one description (90% employment) and other participants with the very same situation but using another description (10% unemployment). Their control over the participants’ information environment allowed them to test the underlying assumption. In this case, the context was compelling not because it matched some “real-world” referent, but rather, because it showed how contextual changes that were thought to be theoretically innocuous do in fact shape people’s expressed preferences.

These findings forced a fundamental and constructive rethinking of the role of formal logic in political explanations, because many formal models make assumptions that were found to violate the experimental results. The experiments also sparked increased interest in the psychological underpinnings of political behavior in fields as diverse as international relations and health policy. Perhaps ironically, one criticism of Quattrone and Tversky’s experiments is that they are not sensitive enough to context. In many political contexts, individuals have access to more than one source of information. Such critiques motivate Druckman’s (2004) experiments and reveal the sensitivity of Quattrone and Tversky’s results to environments where individuals can talk to one another or access competing information.

## Whispering To Princes

Many experimental studies *create* a political context for purposes of testing a causal claim. Field experiments, by contrast, take advantage of *naturally occurring* political contexts while simultaneously leveraging the inferential benefits of random assignment. Because of their “realistic” foundations, field experiments can be especially relevant to policymakers. For example, experimental studies gauging the effectiveness of campaign tactics have changed both scholarly understanding of electoral mobilization and the way in which campaigns allocate their resources.

The first randomized experiment published in the *Review* was Eldersveld’s (1956) study of voter mobilization, which examined the extent to which direct mail, telephone calls, and door-to-door canvassing increased voter turnout in a pair of municipal elections. This path-breaking experiment varied both the content of the mail appeals (“rational” versus “emotional”) and the types of canvassers (party activists versus students). Moreover, it endeavored to measure voter turnout using administrative records rather than subjects’ self-reports. (Figure 2 shows Eldersveld interviewing Eleanor Roosevelt.)

Eldersveld’s experimental approach was revisited a half-century later, when Gerber and Green (2000) conducted a study of voter mobilization using direct mail, telephone calls, and face-to-face visits in an effort to assess whether the long-term decline in voter turnout is causally related to the decline of face-to-face mobilization and the corresponding rise of direct mail and professional phone banks. Like Eldersveld, Gerber and Green measured outcomes using

**FIGURE 2. Samuel Eldersveld interviewing Eleanor Roosevelt**

Courtesy of Michigan Public Media and the Department of Political Science at the University of Michigan.

administrative records rather than self-reports but did so using a sample size more than 100 times as large. Their findings suggest that face-to-face mobilization is indeed more effective in stimulating voter turnout than techniques such as direct mail and phone calls from telemarketing firms, which have largely supplanted it. In the brief period since the publication of this study, dozens of field experiments have been conducted on topics ranging from email communication to television advertising. The assessment of voter mobilization tactics is becoming the most exacting empirical literature in political science, thanks in part to the sheer size and diversity of experimental studies. Gerber and Green (2005), for example, in an effort to replicate and extend disputed findings, conducted a voter mobilization experiment involving more than 1 million subjects.

Such experiments provide an alternative method for studying the question “Do campaigns matter?” When observational studies analyze the correlation between campaign content and voting behavior, they inevitably confront an inference problem: do voters have varying political preferences because they encounter different campaign messages, or do they encounter different campaign messages because their political preferences cause them to be exposed to distinct messages? Experiments solve this conundrum by randomly assigning different types of communication to voters. As a result, the communication to which voters are exposed is statistically independent of the voters’ background attributes. Moreover, these experiments allow for nuanced tests of theoretically interesting phenomena, such as the extent to which externally induced increases in turnout in one election persist in subsequent elections (Gerber, Green, and Shachar 2003) or dif-

fuse through social networks (Nickerson 2006). Scholars have even begun to examine whether randomly generated grassroots activity affects legislative outcomes (Bergan 2006).

Because these field experiments focus on the effectiveness of interventions that are commonly used in politics, they attract attention from those who fund, manage, or regulate campaigns. The substantive findings of these experiments have encouraged a resurgence of interest in grassroots campaign tactics. They have also fueled a collaborative relationship between campaigns and scholars, who share an interest in understanding attitudes and behavior.

## A QUANTITATIVE ANALYSIS OF THE IMPACT OF EXPERIMENTAL ARTICLES

It is one thing that experiments have become a larger part of the discipline; it is quite another whether scholars have paid attention to them. Citation analysis provides a measure of the impact of the *Review*’s experimental articles considered as a whole. It allows us to address the question of whether *Review* articles that employ the experimental method typically receive as much attention as work that uses other approaches.

We used the online Web of Science Social Sciences Citation Index, which reports the number of times an article has been cited since its publication, to obtain the number of citations for each experimental article. We then did the same for a comparable non-experimental *Review* article. Because what constitutes a comparable article is debatable, we examined three alternatives. First, we drew a random sample of approximately six nonexperimental articles in every *Review* volume in which at least one experimental article appeared. This comparison addresses the question of whether experimental articles have, on average, achieved comparable prominence to other contemporaneous *Review* articles. Second, we narrowed this random sample so that it consisted solely of quantitative articles. This comparison speaks to the question of whether experimental articles receive as much attention as articles whose methodological approach and format are more similar. Finally, we further narrowed the range of content by pairing each experiment with two articles on the same general substantive topic that appeared in the *Review* during the same year or the year before.

We found that experimental articles ( $n = 57$ ) were cited an average of 40.1 times ( $SD = 40.5$ ), whereas concurrent nonexperimental articles ( $n = 193$ ) were cited an average of 30.5 times ( $SD = 37.7$ ). Concurrent quantitative articles were cited an average of 29.7 times ( $n = 89$ ,  $SD = 37.0$ ). Approximately concurrent articles that were also matched for content ( $n = 114$ ) were cited an average of 32.8 times ( $SD = 41.3$ ).

For the first two comparisons, citation counts were regressed, in a negative binomial model, on a dummy variable scored 1 if the article is experimental and 0 otherwise. The regression also controlled for year-level fixed effects. When the comparison set is a random sample of non-experimental articles, the estimated effect of the experimental dummy is .38 with a standard error

of .14 (one-tailed  $p = .003$ ). This estimate implies that in any given year, experimental articles have an expected citation rate that is approximately 47% higher than their nonexperimental counterparts. The effect is somewhat stronger when we compare experimental articles to their contemporaneous counterparts ( $b = .55$ ,  $SE = .16$ ,  $p < .001$ ), implying a 74% edge. Comparing experimental articles with articles in the same substantive domain shows a less marked but nonetheless significant edge. With fixed effects for each of the 57 matched comparisons, experimental articles maintain a 26% advantage in citation counts ( $b = .23$ ,  $SE = .12$ ,  $p < .05$ ). Although citation counts are an imperfect measure of scholarly impact, they nevertheless corroborate the qualitative impression that experiments have made enduring contributions.<sup>6</sup>

## THE STRENGTHS, LIMITATIONS, AND FUTURE OF EXPERIMENTS IN POLITICAL SCIENCE

When a research literature evolves to a point where theory, observational studies, or policy concerns generate contested causal claims and when potential problems such as two-way causation and omitted variable bias plague the statistical analysis of observational data, experiments can offer an effective method for adjudicating disputes. The growing presence of experiments in our discipline, however, has not arisen only from an increasing demand to clarify causal questions; it also comes from the changing supply in the experiments themselves. In political science, where context is a defining element, experimental studies are most compelling when they offer transparent tests in relevant environments. Methodological and technological advances have enhanced researchers' capacities to construct novel experimental contexts or to insert experimental manipulations within natural-occurring situations.

The increasing use of experiments thus stems, in part, from innovations that expand the applicability of experiments to new questions and to various times, settings, stimuli, and populations. Creative experimentalists have constructed realistic settings by making the laboratory akin to board rooms and living rooms. They have taken advantage of new computer software, particularly computer-assisted telephone interviewing, to carry out complex studies embedded in surveys, with representative samples;<sup>7</sup> and they have used new metrics such as response latency measures that capture the time it takes to respond to a stimulus. The development of induced value theory in experimental economics

<sup>6</sup> The finding that experimental articles receive *more* attention than other articles should be taken with some caution. As with any emergent method, the threshold for publication in the *Review* may have been relatively high for experimental articles, meaning that those published reflect particularly important studies. Thus, the rate of citations to experimental articles overall may level off over time.

<sup>7</sup> All survey experiments that we identified as appearing in the *Review* relied on non-student samples. In contrast, about 64% of the laboratory experiments used exclusively students, with another 17% using a mix of students and nonstudents.

has facilitated the creation of environments that mimic situations that decision-makers often face in political contexts.<sup>8</sup> And innovative field experiments enable researchers to carry out large-scale studies with direct policy relevance.<sup>9</sup>

Nevertheless, when all is said and done, critics may claim that experimental research is inherently limited to specialized fields in the discipline, especially formal theory, public opinion research, electoral politics, and legislative politics. Most of the experiments published in the *Review's* first 100 years do fall into these categories. So although *Review* experiments have touched on topics as broad as the management of collective action problems (Ostrom, Walker, and Gardner 1992), they have made limited contributions to the study of other topics ranging from civil war to terrorism to regime stability.

But are experiments inherently limited from contributing to such fields? For example, an early and highly influential experimental research program in political science used role-playing simulations to test how various situations affected decisions to go to war (e.g., Guetzkow and Valadez 1981, Mahoney and Druckman 1975). Hermann and Hermann (1967) applied this approach to study the causes of World War I. More recently, Tetlock and Lebow (2001) studied the impact of alternative historical outcomes by asking experts to consider subsets of counterfactual endings to events like the Cuban Missile Crisis; they found that such exposure led the experts to become more inclined to believe in their possibility than those who were not asked. Such examples offer previews of substantive domains in which experiments have some potential to aid inference. Such examples also accentuate the fact that experiments are a methodology that can complement other approaches. Detailed historical work preceded the implementation of the Tetlock and Lebow experiments. Their subsequent experimental results provided a blueprint of what variables might be worth exploring in future studies of foreign policy decision-making and war. Such examples suggest that experiments, and research designs that are influenced by experimental reasoning, have the potential to add value and power to many other areas of inquiry.

Another reason for cautious optimism about the prospect for experimental research is the rapid evolution of experimental reasoning itself. Recent years have witnessed a dramatic increase of experimental research throughout the social sciences. This trend has been accompanied by a growing interest in the statistical underpinnings of causal inference and increasing awareness that even nonexperimental research borrows its logic and assumptions from experimental

<sup>8</sup> Of the laboratory experiments we identified as appearing in the *Review*, half employed induced value theory, such that participants received financial rewards contingent on their performance in the experiment. Of the laboratory experiments, 31% used deception; no experiments used both induced value and deception.

<sup>9</sup> Of the 57 experimental studies we identified in the *Review*, 7% were carried out in naturalistic field settings, 63% took place in simulated settings such as a laboratory or classroom, and 30% occurred in telephone, face-to-face, or mail surveys.

analogs. Equally important is the growing sense, best articulated by Donald Campbell (1969), that experimental investigation should be an integral part of policy innovation, so that society can draw reliable lessons about the consequences of social, political, or economic change. If Campbell's "experimental ethos" continues to grow, subfields in political science that are currently viewed as beyond the purview of experimental investigation might have a stock of experimental knowledge a century from now.

Experiments in political science have progressed from a method used in the occasional anomalous study to a generally accepted and influential approach. This evolution was driven by the maturation of research literatures that demand acute tests of causal claims, and by innovations in the implementation of experiments that expand their reach. Political science experiments can transform—have transformed—thinking on a topic when carried out in relevant contexts—and to be relevant, the situation need not be isomorphic with a naturally occurring (i.e., "real world") referent. Rather, the evaluation of an experiment's power depends on the precise claim under study, the general usage (e.g., searching for facts, speaking to theorists, or whispering to princes), and the specific goal of the study, including its connection to related work that often uses alternative methodologies. Political science might not be an experimental discipline, but with creativity it can become a discipline whose contributions are deepened and strengthened by experimental research.

## REFERENCES

- Ansolabehere, Stephen, Shanto Iyengar, Adam Simon, and Nicholas Valentino. 1994. "Does Attack Advertising Demobilize the Electorate?" *American Political Science Review* 88 (December): 829–38.
- Bartels, Larry M. 1993. "Messages Received." *American Political Science Review* 87 (June): 267–85.
- Bergan, Daniel E. 2006. "Does E-Activism Work?" Unpublished manuscript, Institution for Social and Policy Studies, Yale University.
- Campbell, Donald T. 1969. "Reforms as Experiments." *American Psychologist* 24 (4): 409–29.
- Clarke, Harold D., Allan Kornberg, Chris McIntyre, Petra Bauer-Kaase, and Max Kaase. 1999. "The Effect of Economic Priorities on the Measurement of Value Change." *American Political Science Review* 93 (September): 637–47.
- Cook, Thomas D., and Donald T. Campbell. 1979. *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Boston: Houghton Mifflin Company.
- Cover, Albert D., and Bruce S. Brumberg. 1982. "Baby Books and Ballots." *American Political Science Review* 76 (June): 347–59.
- Druckman, James N. 2004. "Political Preference Formation." *American Political Science Review* 98 (November): 671–86.
- Druckman, James N., and Arthur Lupia. 2006. "Mind, Will, and Choice." In Charles Tilly and Robert E. Goodin, (eds.). *The Oxford Handbook on Contextual Political Analysis*. Oxford: Oxford University Press.
- Druckman, James N., and Joanne M. Miller. 2004. "The Political Psychology of Electoral Campaigns." *Political Psychology* 25 (4): 501–506.
- Eldersveld, Samuel J. 1956. "Experimental Propaganda Techniques and Voting Behavior." *American Political Science Review* 50 (March): 154–65.
- Fiorina, Morris P., and Charles R. Plott. 1978. "Committee Decisions under Majority Rule." *American Political Science Review* 72 (June): 575–98.
- Fisher, Ronald. 1935. *Design of Experiments*. New York: Hafner Publishing.
- Fréchette, Guillaume, John H. Kagel, and Steven F. Lehrer. 2003. "Bargaining in Legislatures." *American Political Science Review* 97 (June): 221–32.
- Gerber, Alan S., and Donald P. Green. 2000. "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout." *American Political Science Review* 94 (September): 653–63.
- Gerber, Alan S., and Donald P. Green. 2005. "Correction to Gerber and Green (2000), Replication of Disputed Findings, and Reply to Imai (2005)." *American Political Science Review* 99 (June): 301–313.
- Gerber, Alan S., Donald P. Green, and Ron Shachar. 2003. "Voting May be Habit Forming." *American Journal of Political Science* 47 (July): 540–50.
- Gilens, Martin. 1996. "'Race Coding' and White Opposition to Welfare." *American Political Science Review* 90 (September): 593–604.
- Green, Donald P. 1992. "The Price Elasticity of Mass Preferences." *American Political Science Review* 86 (March): 128–48.
- Guetzkow, Harold, and Joseph J. Valadez, eds. 1981. *Simulated International Processes*. Beverly Hills, CA: Sage.
- Hermann, Charles F., and Margaret G. Hermann. 1967. "An Attempt to Simulate the Outbreak of World War I." *American Political Science Review* 61 (June): 400–16.
- Inglehart, Ronald. 1971. "The Silent Revolution in Europe." *American Political Science Review* 65 (December): 991–1017.
- Inglehart, Ronald. 1977. *The Silent Revolution: Changing*. Princeton NJ: Princeton University Press.
- Inglehart, Ronald. 1990. *Culture Shift in Advanced Industrial Society*. Princeton NJ: Princeton University Press.
- Iyengar, Shanto, Mark D. Peters, and Donald R. Kinder. 1982. "Experimental Demonstrations of the 'Not-So-Minimal' Consequences of Television News Programs." *American Political Science Review* 76 (December): 848–58.
- Lasswell, Harold D. 1951. "The Immediate Future of Research Policy and Method in Political Science." *American Political Science Review* 45 (March): 133–42.
- Lau, Richard R., and David P. Redlawsk. 1997. "Voting Correctly." *American Political Science Review* 91 (September): 585–98.
- Lodge, Milton, Kathleen M. McGraw, and Patrick Stroh. 1989. "An Impression-driven Model of Candidate Evaluation." *American Political Science Review* 83 (June): 399–419.
- Lodge, Milton, Marco R. Steenbergen, and Shawn Brau. 1995. "The Responsive Voter." *American Political Science Review* 89 (June): 309–26.
- Lowell, A. Lawrence. 1910. "The Physiology of Politics." *American Political Science Review* 4 (February): 1–15.
- Mahoney, Robert, and Daniel Druckman. 1975. "Simulation, Experimentation, and Context." *Simulation & Games* 6 (September): 235–70.
- McDermott, Rose. 2002. "Experimental Methods in Political Science." *Annual Review of Political Science* 5: 31–61.
- McGraw, Kathleen M., and Valerie Hoekstra. 1994. "Experimentation in Political Science." *Research in Micropolitics* 3: 3–29.
- Morton, Rebecca B. 1993. "Incomplete Information and Ideological Explanations of Platform Divergence." *American Political Science Review* 87 (June): 382–92.
- Morton, Rebecca B., and Kenneth C. Williams. 2006. "Experimentation in Political Science." In *The Oxford Handbook of Political Methodology*, ed. Janet Box-Steffensmeier, David Collier, and Henry Brady. Oxford: Oxford University Press.
- Mutz, Diana C. 2002. "Cross-Cutting Social Networks." *American Political Science Review* 96 (March): 111–26.
- Nelson, Thomas E., Rosalee A. Clawson, and Zoe M. Oxley. 1997. "Media Framing of a Civil Liberties Conflict and Its Effect on Tolerance." *American Political Science Review* 91 (September): 567–83.
- Nickerson, David W. 2006. "Measuring Interpersonal Influence." Doctoral thesis, Department of Political Science, Yale University.

- Ostrom, Elinor, James Walker, and Roy Gardner. 1992. "Covenants with and Without a Sword." *American Political Science Review* 86 (June): 404–17.
- Quattrone, George A., and Amos Tversky. 1988. "Contrasting Rational and Psychological Analyses of Political Choice." *American Political Science Review* 82 (September): 719–36.
- Roth, Alvin E. 1995. "Introduction to Experimental Economics." In *The Handbook of Experimental Economics*, ed. John H. Kagel and Alvin E. Roth. Princeton, NJ: Princeton University Press.
- Sniderman, Paul M. 1995. "Evaluation Standards for a Slow-Moving Science." *PS: Political Science and Politics* 28 (September): 464–67.
- Sniderman, Paul M., Richard A. Brody, and Philip E. Tetlock. 1991. *Reasoning and Choice: Explorations in Political Psychology*. Cambridge: Cambridge University Press.
- Sniderman, Paul M., and Douglas B. Grob. 1996. "Innovations in Experimental Design in Attitude Surveys." *Annual Review of Sociology* 22: 377–99.
- Sniderman, Paul M., Look Hagendoorn, and Markus Prior. 2004. "Predispositional Factors and Situational Triggers." *American Political Science Review* 98 (February): 35–50.
- Tetlock, Philip E., and Richard Ned Lebow. 2001. "Poking Counterfactual Holes in Covering Laws." *American Political Science Review* 95 (December): 829–43.