

This version 04/05/2006

## **Experimentation in Political Science**

Rebecca B. Morton  
New York University  
[rbm5@nyu.edu](mailto:rbm5@nyu.edu)

Kenneth C. Williams  
Michigan State University  
[willia59@msu.edu](mailto:willia59@msu.edu)

To appear in *The Oxford Handbook of Political Methodology*  
edited by Janet Box-Steffensmeier, David Collier, and Henry Brady

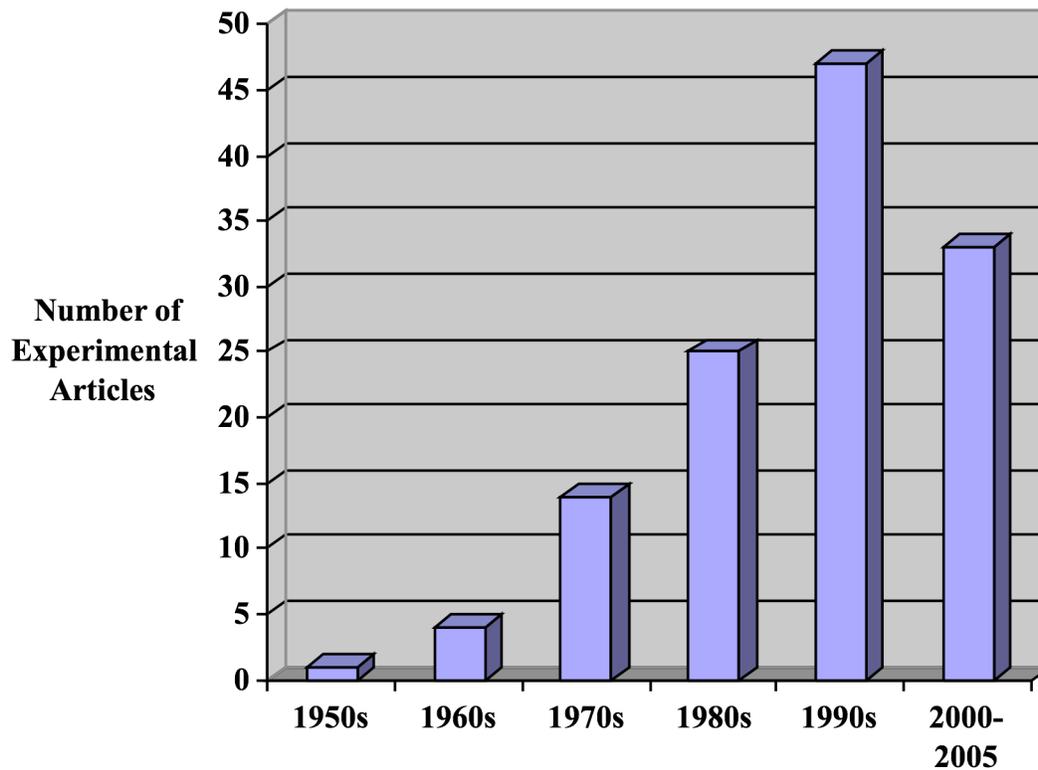
### **The Advent of Experimental Political Science**

Experimentation is increasing dramatically in political science. Figure 1 shows the number of experimental articles published by decade in three major mainstream journals—*American Political Science Review* (APSR), *American Journal of Political Science* (AJPS), and *Journal of Politics* (JOP)—from 1950 to 2005.<sup>1</sup> These figures do not include the new use of so-called “survey experiments” and “natural experiments”. Moreover, other journals have published political experimental articles, such as *Economics and Politics*, *Political Behavior*, *Political Psychology*, *Public Choice*, *Public Opinion Quarterly*, *Journal of Conflict Resolution*, as well as numerous economics and social psychology journals. Furthermore, a number of political scientists have published experimental work in research monographs, see for example Ansolabehere and Iyengar (1997), Lupia and McCubbins (1998), Morton and Williams (2001).

---

<sup>1</sup> Earlier figures are from McGraw and Hoekstra (1994); more recent ones were compiled by the authors.

**Figure 1: Experimental Articles Published in APSR, AJPS, & JOP 1950-2005**



There are a number of reasons for the increase in experimentation in political science in the 1970s and 1980s as is discussed in Kinder and Palfrey (1993). We believe that the dominant explanation for the expansion since 1990 is the increase in cheap and easily programmable computer networking technology. Via such networks in laboratory and over the Internet, the number of possible experimental designs is hugely greater than what researchers can conduct manually. Computer technology has also led to a greater ability to engage in survey experiments and to deal with the statistical and other

methodological issues that are sometimes involved in field and natural experiments. Technology has transformed political science into an experimental discipline. In this chapter we explore the expanding area of experimental political science and some of the concerns involved in experimentation.<sup>2</sup>

### **What is an Experiment?**

#### **The Myth of the Ideal Experiment**

Political science is a discipline that is defined by the substance of what we study, politics. As Beck (2000) remarks, researchers “freely use whatever methodological solutions are available” drawing from other social sciences like economics, sociology, and psychology as well as statistics and applied mathematics to answer substantive questions. Unfortunately this has led to misinterpretations as experimentalists learn about their method from other disciplines and few nonexperimentalists have an understanding of advances that have occurred outside political science. The most prominent misconception is that the ideal experiment is one where a researcher manipulates one variable, called a treatment, there is an experimental group who receives the treatment and a control group who does not, and subjects are randomly assigned across groups. This perception arises largely because most nonexperimentalists learn about experiments through the lens of behavioral social science methods formulated in the 1950s. Certainly such a design can be perfect for the case where a researcher is interested in the causal effect of a binary variable that might affect the choices of subjects in some known context, and the researcher is interested in nothing else. But this is rarely true for any significant question in twenty-first century political science research. Furthermore,

---

<sup>2</sup> A full and detailed treatment of experimentation in political science cannot be presented in a chapter length study and we encourage readers to consult Morton and Williams (2006), which expands on the concepts and issues presented here.

advances in technology, as discussed above, have changed the nature of experimental research in fundamental ways that researchers writing in the 1950s could not have anticipated. By having an outdated view of what an experiment should be and what is possible with experimentation, political scientists often rule out experimentation as a useful method for many interesting research questions.

There is no perfect or true experiment. The appropriate experimental design depends on the research question just as is the case with observational data. In fact, the variety of possible experimental designs and treatments is just as wide as it is with observational data, and in some ways there is a greater range of possibilities through experiments than is possible with observational data. How then is experimental research different from research with observational data? What is experimental political science?

### **Playing God**

The defining characteristic of experimental research is intervention by the researcher in the *data generating process* (DGP), which we define as the source for the data we use in research. In experimental research, the variation in the data is partly a consequence of the researcher's decisions before the data is measured, the design stage. We call the data generated by such intervention, *experimental data*. Non-experimental empirical research involves using only data in which all the variation is a consequence of factors outside of the control of the researcher, the researcher only observes the DGP, but does not intervene in that process.<sup>3</sup> We call this data *observational or non-experimental data*.

---

<sup>3</sup> This assumes of course that the researcher measures the data perfectly; clearly choices made in measurement can result in a type of post DGP intervention.

### **Simulations versus Experiments**

Occasionally political scientists use the term experiment to refer to a computer simulation either to solve a formal model or to explore the implications of an estimated empirical model. We do not include such analyses here. In experimental political science the subjects' choices are independent of the researcher's brain; the subjects' brains and feelings affect their choices in ways that the researcher cannot control. As opposed to simulations, subjects in experiments freely make decisions within the controlled environment constructed by the experimenter.

### **Natural Experiments**

Sometimes nature acts in a way that is close to how a researcher, given the choice, would have intervened. For example, hurricane Katrina displaced thousands of New Orleans residents and changed the political makeup of the city as well as having an impact on locations that received large numbers of refugees. This gives researchers the opportunity to evaluate theories of how representatives respond to changing constituencies. Such situations are called *natural experiments*. It is *as if* nature had two sides, the side that generates most data, and then the interventionist side that occasionally ran experiments like academics, messing up its own data generating process.

### **Aren't We Leaving Important Things Out?**

Beyond intervention, some would contend that control and random assignment are "must have" attributes to an experiment. Below we discuss these two additional features.

## Control

In the out of date view control refers to a baseline treatment that allows a researcher to gather data where he or she has not intervened. But many experiments do not have a clear baseline and in some cases it is not necessary. For example, suppose a researcher is interested in evaluating how voters choose in a three-party election conducted by plurality rule as compared to how they would choose in an identical three-party election conducted via proportional representation. The researcher might conduct two laboratory elections where subjects' payments depend on the outcome of the election but some subjects vote in a proportional representation election and others vote in a plurality rule election. The researcher can then compare voter behavior in the two treatments.

In this experiment and in many like it, the aspect of control that is most important is not the fact that the researcher has a comparison, but that the researcher can control confounding variables such as voter preferences and candidate identities in order to make the comparison meaningful. To make the same causal inference with observational data alone the researcher would need to do two things: (1) rely on statistical methods to control for observable confounding variables such as control functions in regression equations, proximity scores, or matching methods and (2) make the untestable assumption that there are no unobservable variables that confound the causal relationship the researcher is attempting to measure.<sup>4</sup> In observational data things like candidate identities cannot be held constant and it may be impossible to "match" the observable variables. By being able to control these things, an experimentalist can ensure that observable variables are perfectly matched and that many unobservables are actually made observable and subject to the same matching.

---

<sup>4</sup> Morton and Williams (2006) discuss these issues in depth.

### **Random Assignment**

Most non-experimentalists think of random assignment as the principal way that an experimentalist deals with unobservables. If subjects are randomly assigned to treatments, then the experimentalist can eliminate, within statistical limits, extraneous factors that can obscure the effects that the experimentalist expects to observe (or not observe) such as voter cognitive abilities, gender differences in voter behavior, and so on. For this reason, experimentalists typically attempt to randomly assign manipulations as much as possible and when random assignment is not perfect use statistical techniques to account for uncontrolled observables.

## **The Validity of Experimental Political Science Research**

### **Defining Validity**

The essence of the validity question of empirical research is what can we believe about what we learn from the data? Shadish, Cook, and Campbell (2002), hereafter SCC, use the term validity as “the approximate truth” of the inference or knowledge claim. As McGraw and Hoekstra (1994) remark, political scientists have adopted a simplistic view of validity based on the early division of Campbell (1957), using internal validity to refer to how robust experimental results are within the experimental data and external validity to refer to how robust experimental results are outside the experiment. Again, political scientists are using an outdated view. There are actually many ways to measure the validity of empirical research, both experimental and observational. SCC divide validity into four types: *statistical conclusion validity*, *internal validity*, *construct validity*, and *external validity*. Statistical conclusion validity is defined as whether there is a statistically significant covariance between the variables the researcher is interested in

and whether the relationship is sizeable. Internal validity is defined as the determination of whether the relationships the researcher finds within the particular dataset analyzed are causal. This is a much more narrow definition than generally perceived by many political scientists. In fact, Campbell (1986) re-labeled internal validity to local molar causal validity.

When some political scientists think of internal validity, particularly with respect to experiments that evaluate formal models or take place in the laboratory, they are referring to what SCC call construct validity. Construct validity has to do with how valid the inferences of the data are for the theory (or constructs) the researcher is evaluating. SCC emphasize issues of sampling accuracy—is this an appropriate dataset for evaluating the theory? Is there a close match between what the theory is about and what is happening in the manipulated data generating process? Thus, when political scientists refer to internal validity they are often referring to three things as a whole: statistical conclusion validity, local molar causal validity, and construct validity.

### **Are Experiments Externally Valid?**

SCC defines external validity as whether causal inferences established in an empirical analysis hold over variation in persons, settings, treatment variables, and measurement variables. Do the results from one dataset generalize to another? This is a question that is of importance for any empirical researcher, regardless of whether the dataset is experimental or observational. If we study, for example, voter turnout in Congressional elections, can we generalize from that study to turnout in mayoral elections or elections in Germany or elections held in a laboratory? Yet, most political scientists rarely worry about the external validity issue with observational data but do

worry about it with experimental data. Political scientists typically assume that external validity means that the dataset used for the analysis must resemble some “natural” DGP. Since experimentation means manipulation and control of the DGP and an experimental dataset generated through that manipulation and control, then some argue it is obviously not natural and the results cannot be externally valid by definition. Yet even in observational data, the researcher, by choosing variables to measure and study, by focusing on particular aspects of the data, also engages in manipulation and control through statistical and measurement choices. The researcher who works with observational data simplifies and ignores factors or makes assumptions about those things he or she cannot measure. The observational dataset is not natural as well since it is similarly divorced from the natural DGP and by the same reasoning the results cannot be externally valid either.

It is a mistake to equate external validity with whether a given dataset used to establish a particular causal relationship resembles the natural DGP (which can never be accurately measured or observed). Instead, establishing whether a result is externally valid involves replication of the result across a variety of datasets. For example, if a researcher discovered that more informed voters were more likely to turnout in Congressional elections and then found that this is also true in mayoral elections, in elections in Germany, and in an experiment with control and manipulation as well, then we would say the result showed high external validity. The same thing is true about results that originate from experimental analysis. If an experiment demonstrates a particular causal relationship, the establishment of whether that relationship is externally valid is accomplished by examining whether the relationship exists across a range of

experimental and observational datasets and not whether the original experimental dataset resembles a hypothesized version of the natural DGP.

One of the most interesting developments in experimental research has been the influence of laboratory experiments that test the behavioral assumptions in rational choice based models in economics. Although these results are extremely hard to discern in observational data, the robustness of some of the behavioral violations has been established across a wide range of experimental datasets. An entire new field within economics has evolved as a consequence, *behavioral game theory*, with a whole new set of perspectives on understanding economic behavior, something that was almost unthinkable twenty years ago. It is odd that many in political science who disagree with assumptions made in rational choice models are also too quick to dismiss the experiments which test those assumptions as not having external validity, when external validity is really about the robustness of these experiments across different formulations, not about whether the experiment resembles the hypothesized natural DGP.

### **Dimensions of Experimentation in Political Science**

#### **Location: Field, Lab, or Web?**

A dimension that is especially salient among political scientists, particularly non-experimentalists, is location. *Laboratory experiments* are experiments where the subjects are recruited to a common location, the experiment is largely conducted at that location, and the researcher controls almost all aspects in the environment in that location *except* for subject behavior. *Field experiments* are experiments where the researcher's intervention takes place in an environment where the researcher has only limited control beyond the intervention conducted and the relationship between the researcher and the

subject is conducted often through variables outside of the researcher's control. The distinction between laboratory and field experiments is fast becoming more complex as new types of combinations of laboratory and field experiments have developed via the Internet. That is, some of the same treatments that formerly could only be conducted in the laboratory are now conducted via the Internet in a virtual laboratory.

### *Why Laboratories?*

Laboratory experiments provide a researcher with greater control. For example, suppose that a researcher is conducting a decision-making experiment that evaluates the extent that different types of campaign advertisements affect voter preferences. In the laboratory, the researcher can show the ads in a specially designed environment that can be held constant across subjects. In the field, the subjects can be shown the ads in their homes or via the internet, but the researcher loses control over other aspects of the environment in which the ads are watched. If these environmental differences vary systematically with which ads are viewed or other factors that affect the subjects viewing of the ads and these factors are unobservable by the researcher, then the comparison between ads can be more difficult to determine. The advantage of control in the laboratory also extends to group experiments such as the one described above where the researcher is comparing plurality rule with proportional representation.

Second, in the laboratory the researcher has the ability to create environments that do not exist at all in observational data. For example, in the laboratory the researcher can conduct group and individual decision-making experiments on voting mechanisms, which are used very little if at all in any country. Many have proposed that approval voting

would be a better choice mechanism than plurality or proportional representation.<sup>5</sup> In the laboratory a researcher can create and conduct elections using new procedures that it would be difficult to convince jurisdictions to adopt on a random basis as in Bassi (2006).

The third advantage of laboratory experiments is that the researcher can induce a wider range of variation than is possible in the field. For example, suppose a researcher theorizes that voters who choose late in sequential voting elections like presidential primaries learn about candidates' policy positions from horse race information about the outcomes of early voting and this learning affects their choices in the elections. In the field the researcher can randomly provide horse race information to later voters during a presidential primary contest and observe how that affects their choices and information measured via survey. But suppose the researcher also expects that the effect is related to the policy positions of the candidates. In the laboratory the researcher can not only provide information randomly but also randomly vary the candidates' policy positions to see if the effect is robust to this manipulation because the laboratory affords the researcher the opportunity to intervene in hundreds of elections, not just one. Thus the researcher in the laboratory can pick up subtle differences in predictions of theories that the researcher in the field cannot pick up because of the limitation in the number and variety of observations.

Finally, there are still some technologies and manipulations used in experiments that require subjects to go to a laboratory. In particular, social scientists are beginning to use fMRI equipment to measure brain activity as subjects make choices [see Dickson and Scheve (2006)]. Currently such equipment is not portable. Some experiments involve subjects in face-to-face interactions as in some free form deliberation experiments and

---

<sup>5</sup> See for example Brams and Fishburn (1983).

require that subjects be physically in the same space in order to test theories about the effects of such interactions. Such experiments can be difficult if impossible to conduct outside of a laboratory.

*Why the Field or the Internet?*

Field and Internet experiments can potentially be both group or individual, although most currently are individual decision-making experiments. One of the first experiments in political science was Gosnell's (1924) study, which was such a field experiment. By sending postcards with information on voter registration and encouragement to vote to some voters and not others, he manipulated the environment of the subjects. Gosnell did not, however, randomly assign his treatments. More recently Gerber and Green (2000) replicated Gosnell's experiments making use of current understandings of statistical inference. Field experiments do not have the same control over the environment afforded in laboratories. Furthermore, field experiments also raise ethical issues not generally involved in laboratory experiments since many of the individuals affected by the experiment are not aware that they are being manipulated. For example, in Wantchekon's (2003) experiment in Benin, candidates were induced to vary their campaign messages in an examination of the effects of different messages on voter behavior. However, voters were unaware that they were subjects in an experimental manipulation.

Survey experiments are an increasingly popular type of field (and sometimes internet) experiment in political science. In a particularly noteworthy survey experiment, Sullivan, Piereson, and Marcus (1978), the authors compared the question format used by the National Election Studies prior to 1964 and after. Researchers using the data had found that the surveys showed significantly different attitudes towards politics. But at the same

time, the survey questions had changed. Was there really a change in attitudes or did the survey question format change account for the results? Sullivan et al. surveyed the same set of voters but randomly determined which set of questions (pre 1964 or after) each respondent would be asked. The researchers found that indeed the change in the format of the survey did imply that the respondents' had different political attitudes and that the empirical analysis comparing the two time periods was fundamentally flawed.

The most exciting advances in survey experimental design is found in the work of Paul Sniderman with computer-assisted telephone interviewing (CATI), see Sniderman, Brody, and Tetlock (1991). Using computers, telephone interviewers can randomly assign questions to subjects, manipulating the question environment. Recently, through the program National Science Foundation funded program Time-Sharing Experiments in the Social Sciences (TESS), Arthur Lupia at the University of Michigan and Diana Mutz at the University of Pennsylvania have facilitated the advancement of survey experiments in political science, by providing researchers with a large, randomly selected, diverse subject population for such experiments and the infrastructure needed for conducting the experiments with this population.

### **Audience or Goal: Theorists, Empiricists, or Policy Makers?**

Experiments tend to vary in the audience that they address and the goal of the analysis. Experimental political scientists have three types of audiences—theorists, empiricists, and policy-makers.<sup>6</sup>

#### *Theory Testing*

Experiments encompass both tests of well articulated formal theories as well as tests of nonformal theories. A formal theory or model is one where assumptions are

---

<sup>6</sup> We build on distinctions discussed in Davis and Holt (1993) and Roth (1994).

explicitly stated about the DGP and the implications of those assumptions are solved for usually using tools from mathematics. A nonformal theory or model is one where hypotheses are proposed about the DGP, but the assumptions behind the hypotheses are not precisely stated. Fiorina and Plott (1978) published one of the first experimental tests of a formal model of committee decision-making. These experiments were expressly designed to test the theory as formulated. In contrast, the work of Lau and Redlawsk (2001) focuses on evaluating a nonformal theory from political psychology that hypothesizes that voters use cognitive heuristics in decision-making.

How does theory testing of formal models work? In formal models the assumptions are of two types—assumptions about institutional and other exogenous factors such as voting rules, bargaining procedures, preference distributions, and so forth, and assumptions about the behavioral choices of the political actors in the context of these exogenous factors. For example, suppose that a formal model predicts that in plurality rule elections without majority requirements voters are more likely to vote strategically for their second preference in elections with more than two candidates than in plurality rule elections with majority requirements. The model makes assumptions about electoral institutions as well as voter rationality and the predictions are behavioral. An experimentalist can create the electoral institution and pay subjects based on the outcomes of voting that induces a preference ordering over the candidates. An example of such a payment schedule is shown in Table 1 below. Note that the payoffs are received based on who wins, not based on how a voter chose. For example, if subjects who are assigned to be type 1 voted for candidate A and that candidate wins the election,

then that subject would receive \$1 for the election period, but if candidate C wins, even though the subject voted for B, the subject would receive only \$0.10.

**Table 1: An Example Payoff Schedule in a Voting Experiment**

| Voter Types | Winning Candidate |        |        | Number of Voters |
|-------------|-------------------|--------|--------|------------------|
|             | A                 | B      | C      |                  |
| 1           | \$1.00            | \$0.75 | \$0.10 | 3                |
| 2           | \$0.75            | \$1.00 | \$0.10 | 3                |
| 3           | \$0.25            | \$0.25 | \$1.00 | 4                |

Under plurality rule, if all voters voted sincerely for the candidate for whom they receive the highest payoff, then Candidate C would win. But if, for example, voters who are type 1 vote strategically for B while the other voters choose sincerely, B would win. Similarly, if voters who are type 2 vote strategically for A while the other voters vote choose sincerely, A would win. As Myerson and Weber (1993) show formally this voting situation all three of these outcomes are equilibria in pure strategies.

Suppose now that in order for a candidate to win, he or she must receive at least 50% of the vote, that is, there is a majority requirement. If no candidate receives more than 50% of the vote, a run-off election is held between the two top vote receivers. In such a case, then even if everyone votes sincerely in the first round election and C wins that vote, C would have to face either A or B in a run-off (assuming that ties are broken by random draws). In a run-off, either A or B would beat C, so C cannot win. Of course voters of types 1 and 2 might engage in strategic voting in the first round as discussed above and then there would be no need for a run-off. Morton and Rietz (2006) show

formally that when there are majority requirements and voters perceive that there is some probability, albeit perhaps small, that their votes can affect the electoral outcome, the only equilibrium is for voters to vote sincerely in the first stage and let the random draw for second place choose which candidate faces  $C$  in a run-off and then wins the election.

The theory then suggests that strategic voting in our payoff matrix is not likely when there are majority requirements, but possible when not. These predictions depend on assumptions about payoffs, voting rules, and the distribution of voter types as well as voter rationality. The experimenter can create the situation that closely matches the theoretical assumptions about payoffs, voting rules, and the distribution of voter types and then evaluate whether the subjects actually choose as the theory would predict. The results from such an experiment then tell the theorist about the reasonableness of their assumptions about human behavior.

### *Searching for Facts*

Political scientists also use experiments to search for facts. But many see this as a less theory-driven enterprise than economists. Economists typically view searching for facts as beginning with a well articulated formal theory, but varying something experimentally that the theory does not suggest should matter to see if it does. For example, as noted in the voting game in Table 1 above when votes are counted according to plurality rule there are multiple equilibria. Myerson and Weber (1993) conjecture, but do not prove, that campaign contributions and public opinion polls can serve as a coordinating device for types 1 and 2 voters. That is, voter types 1 and 2 would like to coordinate on a common candidate. One way such coordination may be facilitated is by coordinating on whichever candidate is ahead in a public opinion poll or in campaign

contributions prior to the election. However, there is no theoretical basis on which to argue that voters will use polls or contributions in this fashion. As reviewed by Rietz (2003), researchers have used experiments to see if voters do use polls and campaign contributions to coordinate; to evaluate the conjecture. The results show that polls and contributions can serve as a coordination device.

In contrast to the approach of relaxing assumptions of theory or investigating conjectures in the context of a theoretically driven model, some prominent political scientists see searching for facts through experiments as a substitute for theorizing. For example, Gerber and Green (2002) advocate experiments as an opportunity to search for facts with little or no theory. They remark (page 820): “The beauty of experimentation is that one need not possess a complete theoretical model of the phenomenon under study. By assigning ... at random, one guarantees that the only factors that could give to a correlation ... occur by chance alone.” Although they go on to recognize the value of having theoretical underpinnings to research in general as a guide they clearly see this as simply being more efficient, not as a necessary component for experimental searching for facts.

How necessary is a well articulated theory for learning from experiments? Is it possible to discover knowledge without theory? Is theorizing merely an aid to experimental political science, but not a requirement? The argument that searching for facts is possible in an atheoretical setting is based on the presumption that causality can be determined without theory; the claim is that by trial and error of manipulating various causes and determining their effects using experiments, facts and nuances can be accumulated eventually leading to greater knowledge with or without theory. But what

does it mean to measure causality in the experimental setting? Measuring causality *requires* that a researcher think hypothetically or theoretically, have a model of the data generating process.

To see how theorizing is necessary for searching for facts, consider a simple field experiment on voter information. Suppose that a researcher randomly assigns a set of voters in an upcoming election to a treatment group that receives informative campaign material and a control group that does not. The researcher then measures how voters in the treated and control groups voted. In order to make a causal inference about the effect of the treatment on voter behavior, researchers who are searching for facts like Gerber and Green typically take a Rubin Causal Model approach, hereafter RCM, to establishing causality [see Rubin (1974) and Holland (1988)]. The researcher using an RCM approach to establish a causal inference must hypothesize that for each voter in each group there are two potential behavioral choices, a choice the voter makes when informed and a choice a voter makes when uninformed. Of course, only one state of the world is ever observed for any given voter. The researcher must hypothesize that these counterfactual worlds exist even though he or she knows that they do not. Furthermore, the researcher must then make what is called the stable unit treatment assumption (SUTVA) which rules out any cross-effects of treatment from one subject on another's choices, homogeneity of treatment across subjects, and a host of many other implicit assumptions about the DGP that are often unexplored as discussed in Morton and Williams (2006). Finally, the researcher must typically also make the untestable assumption that unobservables do not confound the effect he or she is measuring; that there are no selection effects that interfere with the randomization he or she is using in

the experimental design. Searching for facts, even in field experiments, requires that a researcher theorize; it is not possible to search for facts and through trial and error discover nuances of the DGP without theorizing about that process.

*A Test Bed for Political Methodologists*

Searching for facts is not the only way in which experimental research can provide useful information for empirical scholars. Methodologists often test their methods on simulated data to see if known parameters can be recovered with the method. However, experiments can serve as a *behavioral test bed*. The use of experimental data to test methodologies goes back to at least the work of LaLonde (1986) who compared experimental estimates of the impact of a training program with a range of nonexperimental estimates of the impact of the same program. He found that the nonexperimental estimates varied widely and that although some came close to the experimental estimates there was no a priori way for the empirical researcher to determine which estimates were more reliable. This analysis led to a lot of new developments in the econometric literature on the estimation of treatment effects in observational data as discussed in Morton and Williams (2006).

More recently, Frechette, Kagel, and Morelli (2005) compare empirical procedures used by Ansolabehere, Snyder, Strauss, and Ting (2005) and Warwick and Druckman (2001) to evaluate models of legislative bargaining on data from laboratory experiments where the underlying bargaining framework is controlled by the researchers. Ansolabehere, Snyder, and Ting find support for the Baron-Ferejohn legislative bargaining model [See Baron and Ferejohn (1989)] while, using the same data, Warwick and Druckman (2001) find support for a demand-bargaining model [see Morelli (1999)].

Frechette, Kagel, and Morelli conduct experiments where in some treatments the underlying bargaining model is a Baron-Ferejohn game and in others the underlying bargaining model is a demand-bargaining game. In the laboratory, subjects are paid based on their choices and thus they are induced to have preference as hypothesized by the two theories. Frechette et al. then take the data from each treatment individually and apply the estimation strategies applied by Ansolabehere, et al., and Warwick and Druckman to the experimental data as if they did not know the underlying bargaining game used in that treatment. They find that the empirical estimation strategies fail to distinguish between different bargaining procedures used in the laboratory. Thus, their results show that some standard empirical strategies used by political scientists on observational data unlikely to correctly identify the underlying bargaining game.

*Experimentalists as Politicians' Assistants*

Finally, political scientists also conduct experiments in order to speak to policy makers. Gosnell's early experiment was arguably an attempt to whisper in the ears of princes; it is well known that he was highly involved in everyday politics in his Chicago community. Gerber and Green (2004) have written a book on their experimental research on techniques of mobilizing voters expressly addressed to those involved in election campaigns with a title that is meant to attract practitioners: *Get Out the Vote!*

Moreover, Gerber and Green (2002) have argued that the relationship between political experimentalists in field experiments and actors in politics is a potentially interactive one and that increased experimentation by political scientists will increase the relevance of political science research for political actors. For example, suppose that an interest group is uncertain about which advertising strategy is likely to work in an

upcoming election. If the interest group is induced to randomize their advertisements across television markets by an experimentalist both the experimentalist and the interest group can gain knowledge about the causal effects of such advertisements that presumably the interest group can then use to make more efficient choices in the future.

Should political scientists, as Gerber and Green advocate, increase their focus on experiments that can interest and be directly relevant to nonacademics? Certainly, some experiments can serve this purpose. There is no question that both political scientists and political actors are interested in figuring out why people vote or how campaign advertisements affect voter preferences or how to measure public opinion. Political scientists and political actors are also interested in determining how district boundaries and different electoral rules affect which candidates are elected and in particular, the success of minority candidates. However, as in any research enterprise where researchers focus on questions that mutually interest academics and nonacademics there are potential problems. Given that nonacademics may have financial resources to pay for the services of academics, the weight of their dollars can possibly affect the choice of which questions are asked and how they are asked. Furthermore, nonacademics may not be as willing to share information gathered through such research since politics is often a competitive arena.

### **Concluding Remarks**

Political science is an experimental discipline. Nevertheless, we argue that most political scientists have a view of experimentation that fits a 1950s research methodology and misunderstand the issue of external validity. These false perceptions lead many political scientists to dismiss experimental work as not being relevant to substantive

questions of interest. In this chapter we have provided an analysis of twenty-first century experimentation in political science and demonstrated the wide variety of ways that experimentation can be used to answer interesting research questions and to speak to theorists, empiricists, and policy-makers. We expect experimentation to be an increasingly used tool for political scientists as it has become in other social sciences.

### References

- Ansolabehere, Stephen and Shanto Iyengar. 1997. *Going Negative: How Political Advertisements Shrink & Polarize the Electorate*. New York, NY: Free Press.
- Ansolabehere, Stephen, A. Strauss, James Snyder, and Michael Ting. 2005. "Voting weights and formateur advantages in the formation of coalition governments," *American Journal of Political Science* July 49 (3): 550-563.
- Baron, David and John Ferejohn. 1989. "Bargaining in Legislatures," *American Political Science Review*, 83, 1181-1206.
- Bassi, Anna. 2006. "Experiments on Approval Voting," working paper, New York University.
- Beck, Nathaniel. 2000. "Political Methodology: A Welcoming Discipline," *Journal of the American Statistical Association*, June, vol. 95, no. 450, pages 651-xxx.
- Brams, Steven and Peter Fishburn. 1983. *Approval Voting*. Birkhauser.
- Campbell, Donald T. 1957. "Factors Relevant to the Validity of Experiments in Social Settings," *Psychological Bulletin*, 54:297-312.
- Davis, Douglas and Charles Holt. 1993. *Experimental Economics*, Princeton: Princeton U. Press.
- Dickson, Eric and Ken Scheve. 2006. "", working paper, Yale University.

- Fiorina, Morris and Charles Plott. 1978. "Committee Decisions Under Majority Rule," *American Political Science Review* 72:575-98.
- Frechette, Guillaume, John Kagel, and Massimo Morelli. 2005. "Behavioral Identification in Coalitional Bargaining: An Experimental Analysis of Demand Bargaining and Alternating Offers" *Econometrica* November, 73 (6): 1893-1937.
- Gerber, Alan and Donald Green. 2000. "The Effects of Canvassing, Direct Mail, and Telephone Calls on Voter Turnout: A Field Experiment," *American Political Science Review*, September, 94 (3): 653-663.
- Gerber, Alan and Donald Green. 2002. "Reclaiming the Experimental Tradition in Political Science," in *Political Science: State of the Discipline*, edited by Ira Katznelson and Helen V. Milner, New York: W.W. Norton, pages 805-832.
- Gerber, Alan and Donald Green. 2004. *Get Out the Vote: How to Increase Vote Turnout*, Washington, D.C.: Brookings.
- Gosnell, Harold. 1927. *Getting Out the Vote: An Experiment in the Stimulation of Voting*, Chicago: U. of Chicago Press.
- Holland, P. W. 1988. "Causal Inference, Path Analysis, and Recursive Structural Equation Models" (with discussion), in *Sociological Methodology*, 1988, ed. C. C. Clogg, Washington, DC: American Sociological Association, pp. 449-493.
- Kinder, Donald and Thomas R. Palfrey. 1993. "On Behalf of an Experimental Political Science," in *Experimental Foundations of Political Science*, edited by Donald Kinder and Thomas Palfrey, Ann Arbor: U. of Michigan Press, pages 1-42.
- LaLonde, R. 1986. "Evaluating the Econometric Evaluations of Training Programs." *American Economic Review*, 76, 604-620.

- Lau, Richard R. and David P. Redlawsk. 2001. "Advantages and Disadvantages of Cognitive Heuristics in Political Decision Making," *American Journal of Political Science*, vol. 45, no. 4, October, pages 951-971.
- Lupia, Arthur and Mathew McCubbins. 1998. *The Democratic Dilemma: Can Citizens Learn What They Need to Know?* Cambridge: Cambridge U. Press.
- McGraw, Kathleen and Valerie Hoekstra. 1994. "Experimentation in Political Science: Historical Trends and Future Directions," vol. iv, pp. 3-30, in *Research in Micropolitics*, ed. M. Delli Carpini, Leoni Huddy, and Robert Y. Shapiro, Greenwood, Conn.: JAI Press.
- Morelli, Massimo. 1999. "Demand Competition and Policy Compromise in Legislative Bargaining." *American Political Science Review*. 93, 809-820.
- Morton, Rebecca and Kenneth Williams. 2001. *Learning by Voting: Sequential Choices in Presidential Primaries and Other Elections*, Ann Arbor: U. of Michigan Press.
- Morton, Rebecca and Kenneth Williams. 2006. *Experimental Political Science and the Study of Causality*, unpublished manuscript, New York University.
- Morton, Rebecca and Thomas Rietz. 2006. "Majority Requirements and Strategic Coordination," working paper, New York University.
- Myerson, Roger and Robert Weber. 1993. "A Theory of Voting Equilibria," *American Political Science Review*, March 87 (1): 102-114.
- Rietz, Thomas. 2003. "Three-way Experimental Election Results: Strategic Voting, Coordinated Outcomes and Duverger's Law." Forthcoming in *The Handbook of Experimental Economics Results*, CR Plott and VL Smith, editors, Elsevier Science, Amsterdam.

- Roth, Alvin. 1994. "Introduction to Experimental Economics," in *Handbook of Experimental Economics*, edited by John Kagel and Alvin Roth, Princeton: Princeton U. Press.
- Rubin, D. B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66, 688-701.
- Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Houghton Mifflin.
- Sniderman, Paul M., Richard A. Brody, and Philip E. Tetlock. 1991. *Reasoning and Choice: Explorations in Political Psychology*. New York: Cambridge University Press.
- Sullivan, John L., James E. Piereson, and George E. Marcus. 1978. "Ideological Constraint in the Mass Public: A Methodological Critique and Some New Findings," *American Journal of Political Science* 22:233-49.
- Wantchekon, Leonard. 2003. "Clientelism and Voting Behavior: Evidence from a Field Experiment in Benin," *World Politics*, 55(3):399-xxx.
- Warwick, Paul and James Druckman. 2001. "Portfolio Salience and the Proportionality of Payoffs in Coalition Government," *British Journal of Political Science*, 31, 627-649.