Campaigns as Experiments

Stephen Ansolabehere

EVERY ELECTION YEAR, the traffic rotaries around Boston clog with people carrying brightly colored signs bearing the names of candidates. Kennedy. Kerry. Weld. Menino. Celucci. Swift. The sign bearers come from all walks of life. Some are clad in business suits and look like they took the afternoon off from brokering deals; others wear overalls and union jackets and look like they may have been on the other side of the bargaining table. There are retirees, students, families, and often the candidates themselves. Traffic slows, sometimes to a standstill, as motorists gawk at the spectacle, honking and waving to express their common cause, sometimes gestering to signal their opposition.

This is electioneering in Boston. For those of us who live here, it is an entertaining spectacle and, at times, a nuisance. For visitors caught in a traffic rotary with one hundred screaming bricklayers, Young Republicans, and city workers, it is as strange a ritual as anything Clifford Geertz discovered in New Guinea. Why are these people distracting the already appallingly bad Boston drivers? What could anyone possibly learn about the election from just a sign? Is this what democracy has come to?

While the form may seem unusual, Boston's traffic-circle campaigns raise questions that one may ask of any campaign. What does any of the hoopla that accompanies elections in America signify? There is certainly endless commentary in newspapers and on television programs interpreting the nuances of the competition—the critical importance of the man in the chicken suit who dogged George Bush in 1992, Ed Muskie's and Pat Schroeder's tears, or Gary Hart's 37 percent showing in the New Hampshire primary. Still, the question nags at anyone caught in a traffic circle or deprived of their favorite TV show or just curious. Does any of this matter for who governs and how they govern? Or are campaigns just part of our culture—the stuff of best-sellers and blockbusters?

Driven by the big questions about the relevance of democracy, or perhaps out of a fundamental need to justify our own significance, social scientists have put campaigns under the microscope. In the last half century, we have developed subtle survey research methods that allow us to track the ebbs and flows of public opinion and to tie those to fluctuations in the information that voters receive. We have developed experimental techniques that allow us to isolate the effects of a single bit of information—an advertisement, a news story, an endorsement. We have developed fairly complex statistical methods in order to tease out the subtle relationship between the campaigns and the public's opinions.

A common analytical approach unites this research. Social scientists treat campaigns as experiments. Information—often in the form of advertisements, debates, and conventions—is the *stimulus.* The opinions and the intended and actual behaviors of people are the *responses.* The target of our research is to measure the differences among various treatment groups: those exposed to specific messages and those not exposed. At the grandest level, political scientists treat the entire campaign season as a stimulus, and sometimes democracy itself.

My aim, in this brief essay, is to outline what I see as the methodological challenges that this research program faces. I do not provide a comprehensive survey of the many excellent studies that have been done but instead draw very selectively on research in this area.

The essays in this volume attest to the pervasiveness of experimental thinking about elections and democracy. Nearly every essay in this volume asks what the effects are of different facets of campaign information: the volume of information, the content of information, the timing of information, and the qualities of the candidates who are running. The approach in nearly all of the work here is quasi-experimental, which involves treating our observations of the world as if they were natural experiments. Our discipline has also moved toward more real experiments. The most notable examples are the works of Donald Kinder and Shanto Iyengar; Paul Sniderman and his colleagues; and Charles Plott, Richard McKelvey, and their colleagues. Kinder and Palfrey (1992) present an excellent survey of experimental methods and their applications to political science.

Practitioners more and more treat campaigns as experiments too.

Politicians and their strategists often float ideas and see if they catch on. If a theme resonates with the electorate, it gets more emphasis; if not, it is dropped. The decision about what to use and what not to use is shaped more and more by the tools that social scientists have developed, such as tracking polls and focus groups.

Although experimental reasoning predominates, observation remains the modal approach to the scientific study of elections and democracy. And there is a strong tension between experimentation and observation in social sciences. Unlike physicists and biologists, our "nature" does not perform all possible experiments. Politicians, journalists, consultants, and voters try to anticipate what is the best course and try to avoid actions that might cost them dearly. Consequently, when we look at the world it is always with an eye toward the actions that were not taken, as well as toward the events that did occur. In order to understand to what extent and how campaigns matter, we must imagine worlds with different sorts of campaigns or no campaign at all. This is, I think, one of the greatest challenges for social scientists, and it is one that experimental thinking can help to understand and solve, possibly.

Experimental Reasoning: A Simple Example

To fix what I mean by "experimental thinking," consider a simple example drawn from my research with Shanto Iyengar. We wished to know whether negative advertisements produce lower participation. In other words, we wanted to know whether people who see negative ads are less likely to vote than people who see positive ads. We began by breaking the problem into three components: (1) design of the treatments or stimuli, (2) experimental control, and (3) measurement of responses.

Step 1 requires that we define both the stimulus of interest (exposure to a negative advertisement) and a contrast group. In our study we contrasted exposure to a negative ad with exposure to a positive ad. Call these treatments N (negative) and P (positive).

Skipping ahead to step 3, we wanted to measure whether there were any differences between those who saw a negative ad and those who did not. We measured participatory attitudes many ways using a battery of questions about confidence in government and voting developed by the National Election Studies (NES). We were most interested in intentions to vote; call this dependent or response variable Y. In our final analysis we wanted to compare the responses of those exposed to negative ads (i.e., $Y(\mathrm{N})$) and those exposed to positive ads (i.e., $Y(\mathrm{P})$). The dependent variable Y took a value of 1 if the person intended to vote and 0 if not. The average value of Y (call it y) for each group, then, equaled the percent of people in each group, that is, $y(\mathrm{P})$ and $y(\mathrm{N})$, that intended to vote. The estimated effect of negativity on participation was simply the differences between the average vote intentions of these two groups, that is, $y(\mathrm{P}) - y(\mathrm{N})$. This is just the definition of a difference of means, sometimes casually called an effect. It means something more to say that the effect is causal.

The power of experimental reasoning comes from control. Control allows the researchers to conclude that any statistically important differences in responses to the two treatments are due solely to the experimental manipulation. In the simple example here, we grappled with two problems of control: what do people see (the treatments) and who should see what?

The choice of the two treatments P and N was fundamentally an issue of control. We could contrast (and actually did) the negative ads with a neutral ad, a nonpolitical ad. The problem with such a contrast is that there are actually many differences between the neutral and negative ads: one is about politics and the other is not; one comes from a specific candidate and the other does not; one raises a public policy issue and the other does not; one is negative and the other is not. To conclude unambiguously that it is the negativity of the message that turns people off, we had to eliminate those differences. So we created a positive version of the same ad, using the same video, the same candidate, and the same issue, but changed the negative words in the script to positive ones. In the end, we could be confident that the only difference between the treatments stemmed from the negativity of the message. One could also imagine more elaborate experiments that vary the lengths of the messages or that combine negative and positive messages to make even more refined measures of the effects of the amount of negative and positive information.

The second place where we exerted experimental control was in who saw what ad. More generally, this is an issue of sample selection. If the assignment of people to treatments is at all related to the dependent variable, then there will be biases in the estimated effect of the treatment on the dependent variable. The experimental approach allowed us to measure precisely who watched a negative ad and who watched a positive ad. This is a very important advantage. Surveys and other techniques must rely on reported or recalled exposure to ads. As discussed later, those measures can be highly inaccurate and produce serious biases in data analyses.

At this point in the experiments we introduced a further guard against other factors confounding our conclusion. We randomly assigned advertisements to participants. Another significant problem with simply measuring the difference between those who saw a negative ad and those who saw a positive ad is that the contrast might reflect who is in each group. We might, for example, let people choose which ad they wanted to see. If people sorted such that likely voters chose tape A and likely non-voters chose tape B, then there would be a bias toward finding no effect, even where an effect is present. If the opposite occurred, and people unlikely to vote a priori chose tape A and likely voters chose B, then we would overestimate the effect of negativity. Similarly, almost any assignment rule might introduce biases. To prevent some unanticipated confounding factor from emerging, we randomly assigned (using a random number generator, not an arbitrary rule) tape numbers for participant identification numbers.

Randomization does not remove the features of individual participants that can create biases, such as their inherent likelihood of voting or not or their taste for certain kinds of advertisements. Rather, it provides a statistical safeguard against those dispositions generating biases in the comparison of the two experimental groups. Randomization allows researchers to state that we expect in the statistical sense that any differences between the groups likely did not arise by chance, even though we did not control for their possible causes in the initial design of the experiment. The advantages of control are equally significant. Control allows researchers to say that some specific causes could not have accounted for any differences between the experimental groups that might have emerged. In addition, experimental control is a precondition of randomization. Without control over who sees what, randomization is not possible.

A simple algebraic example helps to clarify these two key advantages to experimentation. This example is based on a very insightful paper by Donald Rubin (1974). Suppose again that we wish to measure the effect of negative ads on participation. Suppose also that we have two individuals we would like to examine, person i and person j. Many factors that affect political behavior will shape these individuals' responses to our study. Indeed, there are so many factors that we could not possibly design an experiment that would hold every one of them fixed through experimental control. What Rubin elegantly explains is how a controlled and randomized experiment can *eliminate from statistical conclusions* the influence of the myriad factors that shape behavior.

To begin with, it is important to understand that complete experimental control is an impossibility. Complete experimental control would mean that we show person i the same two treatments, P and N, and that we show person j the same two treatments, under exactly the same conditions. Suppose that we could set up the experiment so that the conditions of each showing of the videotapes are identical. Index the showings with the letter *t*. We would then measure the difference in person i's participatory intentions between the two treatment conditions and the difference in person j's intentions. Finally, we could average the two viewers' responses to measure what Rubin calls an Average Causal Effect:

$$Y(\mathrm{P}) - Y(\mathrm{N}) = (1/2)\ \{[Y_{it}(\mathrm{P}) - Y_{it}(\mathrm{N})] + [Y_{jt}(\mathrm{P}) - Y_{jt}(\mathrm{N})]\}.$$

It is physically impossible to observe this quantity. An experimenter may be able to set up very similar viewing conditions for the two persons, holding day, time, and so forth, the same. But an individual can never be exposed to two different treatments at exactly the same time. At this point, a solution to the problem of control seems hopeless.

Controlled and randomized experiments allow us to get around this impossibility. Instead of trying to see everything at once, we construct some alternative (counterfactual) situations and then construct a rule to decide which situation we will observe. Continuing with our example, we begin by constructing two hypothetical situations. In situation 1, we determine that person i watches tape P and person j watches tape N. In situation 2, we determine that person j watches tape P and person i watches tape N. We cannot, I have said, observe both of these situations.

The beauty and power of experimentation come at the next step. The experimenter plays God and determines which of these two situations will occur. We may choose which world will occur on the basis of any arbitrary rule. However, there is a very strong reason for selecting by random assignment, such as a coin toss. If we use an arbitrary rule, such as which person comes through the door first, we might inadvertently introduce another reason that any observed differences exist. With the coin toss we know that we can expect that any differences that emerge are due to chance variation, not to systematic differences.

In the advertising experiments, suppose that we tossed a fair coin to determine which of these situations we should observe. So, with probability 1/2 we observe the outcome of situation 1 and with probability 1/2 we observe the outcome of situation 2. Whichever we observe we take to be the estimate of the average causal effect. The key result is that the expected value of the experimentally derived estimate of the effect of advertising tone equals

$$E[y(P) - y(N)] = (1/2)\,[Y_{it}(P) - Y_{jt}(N)] + (1/2)\,[Y_{jt}(P) - Y_{it}(N)].$$

Collecting the terms with P and the terms with N reveals that this quantity is identical to the Average Causal Effect: $Y(P) - Y(N)$. In statistical parlance, the observed outcome of the experiment provides an unbiased estimate of the Average Causal Effect.[1]

Five Challenges in the Study of Campaigns

The structure of a simple experiment helps to clarify the challenges that confront researchers who wish to measure the effects of political information using observational data. There are five: the size of the study, the construction of treatments (for example, P and N), the measurement of exposure (who sees what), the assignment of treatments (why did people see what they saw), and the measurement of responses.

Setting Our Sights

What are the possible effects of campaigns on elections and government? The campaign period is not the only time when people learn about their government, and during the campaigns the public discussions and mass media messages are not the only information that people draw on. Rather, campaigns are just one of the ways that people relate to their government, and this fact must temper our assessment of them.

Often journalists prime us to believe that the latest news is so very important. That is their job: to sell the news. That is also why commentaries in the popular press are so often the starting point of social science inquiries. Although journalists are some of the major consumers of what we discover, it is probably best that we get away from journalistic beliefs about the importance of campaigns, at least as the starting point of any inquiry.

Instead, in thinking about how to study campaigns and whether their effects are big or small, we must determine what a big effect would be in political terms. To my thinking, moving the public opinion polls by 5 to 10 percentage points, which is perhaps an optimistic estimate of the edge gained by congressional incumbents through their campaigns, is a very big effect. Campaigns probably matter a whole lot less than fluctuations in the economy or the ideologies of the parties and people's preferences about the size and extent of government activity. This does not mean that campaigns are unimportant and that we should direct our research elsewhere. Unlike people's preferences about the size of government or fluctuations in the economy, candidates control their campaign messages. Here is the politics of democracy.

The lesson for researchers, though, is that it is hard to measure the effects of campaigns with much confidence. To reliably measure many modest campaign effects at once and the subtle interactions among them requires an excessively large sample. For example, the postconvention bounce enjoyed by the party's standard-bearer is one of the strongest and most reliable campaign effects. The convention bounce averages 6 percentage points. One needs a sample of at least twelve hundred to reliably detect such an effect. Most media polls have samples of five hundred to one thousand. It is harder still to measure the size of effects of this size within particular subgroups, such as among people who identify with one party or another.

What is needed in the study of campaigns is a very focused approach that attempts to isolate a phenomenon of particular importance and then designs the instrument around it. In this volume, I think the best example is the study reported by Johnston and Vowles of strategic voting in New Zealand.

Defining Treatments

Campaigns are important because they are times of intense, focused public discourse. Politicians, parties, and interest groups present the choices to the electorate. For their part, many citizens tune in because they must make a judgment on Election Day about who should govern. Of course, intense public discourse arises at other times, such as the debates over national health insurance and over NAFTA during the first Clinton administration. Indeed, campaigns should be viewed as a continuation of the dialogues that occur throughout the duration of a government. For social scientists, though, they are an ideal time to measure the effects of information because the volume and complexity of the information increase during campaigns and because many more people seek information.

That discourse and information are the real concerns with campaigns suggests that there are two features of campaigns that we may analyze: what is said and how it is said.

The "what" of campaigns are facts. Political scientists, drawing on their Progressive roots, want people to cast informed votes, and our standards of information are high. Voters certainly use a wide range of facts, even if they do not always seem to do so. Indeed, not having heard anything about the candidates is a "fact." A crude, workable typology of facts undergirds most communications and elections research. First, voters need basic information about the election: when is the election and how do you vote? Second, voters use labels, including party, incumbency, group endorsements, race, and gender. Such facts seem necessary for some of the basic sorts of voting that social scientists have shown to exist, especially party voting, retrospective economic voting, incumbency voting, and racial bloc voting. Third, facts about the candidates as persons are also important. Do I know this person? Where does he or she come from? Finally, information about the issues is vitally important to many voters. The twentieth-century ideal of democracy is one based on deliberation about "the issues" facing the country. At the very least, people must have some familiarity with those issues, either at a personal level or at the rarefied level of public debates about specific laws. What is at stake in this election? What are the problems that the nation faces or that I face? What policies do each of the candidates and parties plan to pursue? What might be the consequences of those policies?[2]

To assess the effects of different sorts of information we must draw contrasts between situations where the information is available and where it is not. For example, most ballots are partisan ballots—they have the labels of the parties next to the candidates' names. This bit of information (a simple label provided in the campaign booth) matters quite a lot. I, for one, would be lost without it when voting for lesser offices like city council, sheriff, and judge. The extent to which these labels matter, though, can only be measured by contrasting elections where the labels are present and where they are not. One way to capture such effects is to contrast states with partisan and nonpartisan ballots. Another possible experiment is to consider states that change their ballot form. Minnesota in the early 1960s, for example, experimented with nonpartisan ballots. One can measure the effects of ballot form in this case by contrasting the partisan vote in years when the nonpartisan ballot was used with the partisan vote in the years when the partisan ballot was used.

Significant events, such as debates, conventions, and even the campaign season itself, are often viewed as treatments. Gelman and King (1993), for example, measure a variety of campaign effects, such as convention bounces, by pooling the media polls in the 1988 U.S. presidential election and then testing for the effects of events deemed important by the press coverage of that election. In a similar vein, Thomas Holbrook (1994) measures the effects of the presidential campaigns of the 1980s on election outcomes by contrasting the average presidential popularity and support in the months before the campaign began with the average presidential popularity and support registered throughout the campaign season. This approach gives us a crude assessment of the campaigns—did any movement in opinion occur?

There is an important limitation to such studies. The "treatments" or events may not be comparable. One party's convention might convey less information than the other party's convention. For example, the 1968 Democratic convention is often described as showing a party in disarray, but no one remembers the 1968 Republican convention, nor did it attract nearly as much commentary at the time. One way to adjust for comparability of events is to measure the strength of the signal with television ratings for conventions, advertisements, and other "events." To continue the convention example, researchers might count the number of minutes of issue discussion that occurred during prime time. Even still, it is hard to assay an entire campaign season. How we measure the "treatments" at the aggregate level remains a tricky problem, with few obvious solutions.

In studying the "what" of campaigns, the typology of basic facts, labels, personal traits, and issues offers a primitive but robust guide. But we need better measures of these treatments, especially at the aggregate level. Most of the study of campaigns is conducted at the micro level, largely with survey data. Survey data suffer from measurement errors, as noted later. We need a check on conclusions drawn from surveys, and, short of doing experiments, that means aggregate data. The only standard aggregate data that are available measure the volume of the campaigns, and these are the data reported by the Federal Elections Commission and state agencies on the candidates' campaign expenditures. There are no measures of candidates' expenditures for themselves and against others or of expenditures on different issues or traits.

The "how" of campaigns opens a much wider field for inquiry. By "how" I mean how campaigns are conducted. How many candidates and parties participate? Who controls access to the mass media? At their core, campaigns involve deliberation. Campaigns are protracted events in which the parties, candidates, and electorate attempt to engage in a discourse about what is the best direction for the future government. Indeed, one of the fundamental assumptions of democracy is that there is value in hearing many voices and in interacting among them. This theme was central to the research on political communication following World War II. Control of the mass media seemed necessary for the sustenance of the regimes in Germany and Italy. Social scientists have worked extensively to measure the effects of specific facts and events on opinions and election outcomes. However, we know very little about how the dialogue of campaigns works.

The extent of deliberation has reasserted itself in contemporary research in three important areas. The literature on transitions to democracy has emphasized telecommunications reform as one of the focal points for democratic reforms. The literature on presidential power has focused on the ability of the president to get his way in the Washington establishment by "going public" and on the limitations of this power when Congress is also able to go public (Kernell 1986; Brody 1992). Finally, the popular critiques of the American media and proposals for reform of it often focus on the effects of concentration of ownership and of commercialization on the quality of discourse. I know of no studies that establish such a link in any objective way, but the BBC is often heralded as the model for dragging U.S. politics out of the muck.

While questions of access to media may seem far removed from campaign politics, they are in fact quite central. In 1996, the U.S. Presidential Debate Commission chose to exclude H. Ross Perot from the nationally televised debates. Similar debates four years earlier reinvigorated his campaign and helped him win nearly 20 percent of the vote. In 1998, Reform Party candidate Jesse Ventura was allowed to participate in the Minnesota gubernatorial debate, in part because Democratic nominee Skip Humphrey felt that Ventura (with only 11 percent in the polls) had no shot of winning and would only draw support from the Republican nominee, Norm Coleman. Ventura was the clear winner in the debates and, one month later, in the election.

The number of voices is certainly an important feature of deliberation. Beyond that, though, communications research still lacks an adequate framework for thinking about deliberation and measuring its effects. Three models of deliberative democracy in campaigns have currency. First, one might view voters as jurors, weighing the evidence laid before them by adversaries. Here the normative standards are fairness and truthfulness, and at times an arbiter must intervene. Often, journalists take the role of judge. Second, one might view deliberative democracy as a marketplace. Free and open competition (in our case for elected office) is the best test of any idea or politician. This rationale lies behind many of the Supreme Court opinions relating to campaign finance, broadcasting regulation, and censorship. Finally, one might think of democracy as an extensive town meeting, where ordinary citizens can openly put questions before their fellow citizens, criticize or praise their government, and challenge leaders directly.

How we measure deliberation and discourse poses an even greater research challenge. The very dimensions of the subject have yet to be distinctly defined. Perhaps the best starting point is the vantage of comparative politics, and that is to measure the openness of the media in different societies.

Complexity of Effects

One of the clearest lessons of media research over the last two decades has been the complexity and variety of responses that any message can elicit. Much early social science research on communications emphasized that modern democracy involved the behavior of the masses. Fear drove this research: fear that democracy could fall to the machinations of demagogic dictators. Such fears proved unfounded, mainly because the public is not a single entity easily moved in one direction or another, but consists of many publics.

Any message may have varied effects on the electorate as a whole. Some information may lead to convergence in people's beliefs and voting behaviors, but some information may create greater heterogeneity. Responses to campaign messages vary along at least three significant dimensions.

First, voters differ in their levels of sophistication and knowledge of politics, and this mutes the effects of new information on the public as a whole. More sophisticated voters are more attuned to politics, more likely to pick up new information, and more critical in their use of such information. John Zaller, in his book *The Nature and Origins of Mass Opinion* (1992), develops a parsimonious model that describes how new information can have very uneven effects in such an electorate. A very sophisticated voter is likely to learn new information but is unlikely to be influenced by it, since the new information adds little to the voter's existing cache of knowledge. An unsophisticated voter is very likely to be influenced by new information but is very unlikely to learn it. Voters with moderate levels of information—those who follow public affairs occasionally—are the most susceptible to new messages.

Second, voters differ in their political dispositions and preferences, and people's preferences can shape what information they seek, believe, and respond to. Paul Lazarsfeld and his colleagues (Berelson, Lazarsfeld, and McPhee 1954) discovered in their studies of Elmira, New York, that people learned about politics very selectively. For example, if an individual was unemployed, he or she sought information about how the candidates would get the economy moving again. A politician could sway such a voter by talking about the economy, but the candidates could not sway voters by talking about something else or by promising policies markedly against the individual's interests. Taking party and ideology as manifestations of people's underlying preferences, Shanto Iyengar and I (Ansolabehere and Iyengar 1995a) found a very similar pattern in television viewers' responses to advertising. People were most receptive to candidates who talked about the issues most important to them. They were also most receptive to politicians who were of their same party and who were thus likely to take actions that they would most approve of.

Third, voters face coordination problems, which they at times may be able to overcome. Rather than waste their votes on the candidate who they prefer best but who will certainly lose, voters may switch to their second choice in order to avoid the worst-case candidate. Voters may even choose not to collect information about candidates who have no chance of winning. Calculations of wasted votes will tend to favor moderate candidates, though not always. There is considerable debate in British political science about the extent of sophisticated or strategic voting, which in the 1980s may have hurt the Liberal-Democratic/Alliance party. There are many related phenomena in different electoral systems, such as bandwagons in American presidential primaries and momentum gained from coalition partners in New Zealand (see Johnston and Vowles, this volume; see also Cox 1996).

Social scientists have developed good models for each of these forms of behavior. The challenge for empirical research is one of scale. The number of possible campaign effects is staggering, as the electorate can vary along each of these dimensions, and these dimensions might even have interactive effects. To examine any one of these effects adequately or to detect important interactions among them requires studies that are focused on measuring the effects of campaigns and carefully designed to measure these effects. Only a handful of such studies have really ever been executed.

Measurement of Treatment Exposure

Perhaps the least appreciated problem in the study of campaigns is measuring actual exposure to a message. The great advantage of experiments is that the experimenter observes (and controls) who sees what message. Studies using survey and aggregate data do not measure this directly. In surveys, we may ask whether someone recalled seeing an ad or a story or whether they regularly watch certain programs. In aggregate data analysis, we know the dates that events happen on and can associate those with the time trend in the public opinion polls. These measures do not capture actual exposure.

Determining who actually saw or heard a message has proven a very thorny problem. Schuman and Presser's (1981) novel question-wording experiments show how fragile survey responses can be to slight changes in wording, timing of questions, and sensitivity of the subject. They conclude their research with a preference for using open-ended and uncoached or prefaced questions. In media research, even these forms of questions fail to get at actual exposure.

In our own study, Shanto Iyengar and I tested the advertising exposure question used by NES (Ansolabehere and Iyengar 1995b). About a half hour following the viewing of the videotape we asked participants, "Do you recall seeing any political commercials during the video? If so could you briefly describe the ad?" Using a generous coding, just over half (55 percent) of the people who actually saw an ad could remember that they did. If we required that they could say anything about the ad, that fraction fell to just above a quarter (28 percent). This is a very severe downward bias in the actual exposure rate reported by such a question.

To make matters worse, we found that recall did not mediate the effects of actual exposure. We measured the effects of actual exposure on vote preferences using an ordered probit predicting party preference in the vote on the party of the candidate whose ad was seen, plus many control variables. The effect of the ad was to move vote intentions about 7 percentage points toward the sponsor of the ad. We then measured the effect of recall on vote intentions. People who recalled the ad were only 2 percentage points more likely to vote for the sponsor than those who did not, an insignificant effect. The difference could be due to measurement error, which would bias the coefficient on recall downward, or to a significant mediating effect of recall, which would mean that there is a significant interaction between recalled and actual exposure.

The culprit is measurement error. As the third step in this analysis, we broke the treatment variable into two groups: those who were exposed to and recalled the ad and those who were exposed to the ad and did not recall it. The coefficients were nearly exactly the same. Those who recalled the ad were as strongly affected as those who did not recall the ad. In other words, recall seems to be nothing more than a very bad measure of actual exposure, which strongly influences opinions.

What to do with a measure like this? Perhaps the question should be discarded. Before doing so, though, researchers need to examine whether a valid correction for these measurement errors can be constructed. Are valid instruments available? Can multiple measures fix these problems? The work of Achen (1978) on representation and Bartels (1993) on media exposure generally seem like promising starting points.

Assignment of Treatments

The granddaddy of all media studies problems is the assignment of treatments. When we conduct experiments, randomization allows us to estimate the effects of a specific treatment without bias, eliminating statistically the effects of the many other factors that influence behavior.[3]

In a campaign, what is said and how it is said are dictated by the logic of political strategy, not by the roll of the experimenters' dice. Strategic behavior of candidates and other players will introduce bias into any study if the choice of strategy depends on the expected effect that such actions might have on the vote. For example, in a presidential race, the mathematics of the Electoral College lead candidates to focus their campaign efforts on the swing states. Heavily Democratic states like Massachusetts, New York, and Rhode Island and heavily Republican states like Utah, Alaska, Kansas, and Idaho will see few if any presidential advertisements, while the swing states of Illinois, Michigan, Florida, and Ohio will be inundated with campaign commercials.

The problem this creates for researchers can be seen by considering the hypothetical experiment sketched previously. Now we will choose who sees what by tossing a weighted coin rather than a fair coin. The weight on the coin will be determined by the likely response. Suppose that in the pretest questionnaire we determine that person *i* is more likely to participate than person *j*. In our mathematical symbols given previously, this means that $Y_{it}(P) > Y_{jt}(P)$ and $Y_{jt}(N) > Y_{it}(N)$. Now let us toss a weighted coin where the weights are such that the probability that *i* sees tape P is $q > 1/2$. The expected outcome of the equation now becomes

$$q\,[Y_{it}(P) - Y_{jt}(N)] + (1 - q)\,[Y_{jt}(P) - Y_{it}(N)].$$

This is larger than the Average Causal Effect. If the weight made *j* more likely to see the positive ad, then the estimated effect would be too small.

Of course, if we knew the value of *q* we could fix this quantity with the appropriate weights. Unfortunately, in survey research and analyses of aggregate data, we do not know this quantity. It depends on the behavior of voters, politicians, journalists, and others who produce and demand political information. If we are to reduce the biases that come from nonrandom treatment assignments, we must try to analyze the process that determines who sees what in politics. Three approaches have been used.

First, we may model the cognitive process described earlier. This is another take on John Zaller's model of mass opinion. In his formulation, the probability that someone is exposed to a message is an increasing function of his or her attentiveness to politics and the responsiveness to new information is a decreasing function of his or her attentiveness to politics. Zaller assumes that if a person doesn't receive a message he or she can't be influenced by it, and this allows him to estimate *q* as a function of attentiveness and *Y* as a function of attentiveness. The basic structure of this model can be applied to many other problems, such as the content of messages.

Second, we may use a conventional psychometric solution, which involves multiple measures for each individual and assumptions about the cumulative effects of those measures over time. Bartels (1993) uses such a model to estimate the effects of media exposure, measured using the NES panel data, on opinion formation. Bartels finds modest effects where negative or no correlations existed before. Achen (1983), though, documents the sensitivity of these models to assumptions. Looking at the Miller-Stokes representation data, Achen shows that one model converts correlations between candidates and voters preferences that are in the range of .4 to over .9. But another, equally plausible model pushes those correlations down to .10 or .05. I suspect that, in the end, these psychometric techniques using cross-sectional data or short panels will not prove terribly useful. Estimates are not very robust to specification assumptions, and testing assumptions is extremely difficult and usually impossible. Zaller's approach seems more fruitful, as does the third approach.

Third, we may use the conventional econometric solution of instrumental variables. Instrumental variables estimation requires that researchers measure variables that influence media exposure but not directly political behavior, such as turnout or vote preference. Within political science, these methods have been most widely applied to the study of campaign spending in congressional elections, and there is considerable debate over which sets of variables can be used to make valid instruments (see, e.g., Jacobson 1990; Gerber 1992). The sorts of variables that likely work for campaign spending are factors that affect the cost of raising campaign money, such as the willingness of interest groups to give to members on valuable committees, but do not affect other electoral advantages that incumbents possess. Research on campaign finance is unique in the development of instrumental variables in the study of campaigns, and this approach has considerable promise for other subjects. Needed, though, are systematic measures of factors that affect the volume and content of media coverage of politics.

Conclusions

Strange though many campaign practices may seem, political science has taken a decidedly nonanthropological approach to the study of campaigns. Instead, our field has been informed more and more by the rigors of experimental thinking. We understand the politics of campaigns as causes and effects rather than as its many cultures.

I have sketched, and this essay is surely just a sketch, the main tenets of experimental thinking in the study of campaigns. The strength of experiments is that they offer the most powerful way to observe how the world works under alternative scenarios; they allow us to compare the counterfactuals. Real experiments, though, are relatively rare and often not feasible. Even still, the logic of experimentation clarifies the difficulties of other ways of seeing the world. Analyses of surveys and aggregate data are ideal for mapping the contours of the political landscape. However, when we use surveys, aggregates, and simple observation to measure the effects of information on behavior, we immediately encounter the limitations outlined here. Fixing these problems is difficult but not impossible. And many of the essays in this volume offer innovative attempts to overcome these difficulties.

NOTES

1. With a weighted coin, we would devise an unbiased estimate by using the weights appropriately in the formula given previously.

2. It is tempting to throw many different ideas into this category of facts, including emotions and evaluations of candidates. These are not themselves facts. They are outcomes of the process, and calling them facts fundamentally muddies the enterprise.

3. This problem has long been extensively studied by econometricians and statisticians (Heckman 1978; Imbens, Angrist, and Rubin 1996).

REFERENCES

Achen, Christopher. 1978. "Measuring Representation." *American Journal of Political Science* 22:475–510.

———. 1983. "Toward Theories of Data." In *Political Science: The State of the Discipline,* ed. Ada Finifter. Washington, DC: American Political Science Association.

Ansolabehere, Stephen, and Shanto Iyengar. 1995a. *Going Negative: How Political Advertisements Shrink and Polarize the Electorate.* New York: Free Press.

———. 1995b. "Messages Forgotten." In *Campaigns and Elections American Style,* ed. Candice Nelson and James Thurber. Boulder: Westview.

Bartels, Larry. 1993. "Messages Received: The Political Impact of Media Exposure." *American Political Science Review* 87:267–85.

Berelson, Bernard R., Paul F. Lazarsfeld, and William N. McPhee. 1954. *Voting: A Study of Opinion Formation in a Presidential Campaign.* Chicago: University of Chicago Press.

Brody, Richard, 1992. *Assessing the President.* Palo Alto: Stanford University Press.

Cox, Gary. 1996. *Making Votes Count.* New York: Cambridge University Press.

Gelman, Andrew, and Gary King. 1993. "Why Are American Presidential Election Polls So Variable When Votes Are So Predictable?" *British Journal of Political Science* 23:409–51.

Gerber, Alan. 1992. "Measuring the Effects of Spending in U.S. Senate Elections." Manuscript, Department of Political Science, Yale University.

Heckman, James. 1978. "Dummy Endogenous Variables in a Simultaneous Equation System." *Econometrica* 96:931–59.

Holbrook, Thomas. 1994. "Campaigns, National Conditions, and U.S. Politics." *American Journal of Political Science* 38:973–98.

Imbens, Guido, Joshua Angrist, and Donald Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of Econometrics* 71:145–60.

Jacobson, Gary. 1990. "The Effects of Campaign Spending in House Elections: New Evidence for Old Arguments." *American Journal of Political Science* 34:334–62.

Kernell, Samuel. 1986. *Going Public: New Strategies of Presidential Leadership.* Washington, DC: Congressional Quarterly Press.

Kinder, Donald, and Thomas Palfrey. 1992. *Experimental Foundations of Political Science.* Ann Arbor: University of Michigan Press.

Rubin, Donald. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66:688–701.

Schuman, Howard, and Stanley Presser. 1981. *Questions and Answers in Attitude Surveys.* New York: Academic Press.

Zaller, John. 1992. *The Nature and Origins of Mass Opinion.* New York: Cambridge University Press.