Experimental Designs for Political Communication Research:

From Shopping Malls to the Internet

Shanto Iyengar
Stanford University

Prepared for the Workshop in Experimental Methods
Department of Government, Harvard University, May 5-6, 2000

Contents

Introduction

The Issue of Causal Inference
- Negativity and Candidate Reputation in Campaign Advertising
- Racial Cues in Local News Coverage of Crime
Introduction

Twenty years ago, the use of experimental methods was virtually unknown to the discipline of political science. In the early 1980s, scholarly interest in the interdisciplinary area of political psychology intensified and from there experimental methods began to find their way into several sub-fields of political science, including, gradually, political communication. Despite this methodological transfusion, the field of political communication is still dominated by the use of survey techniques; in the presentation that follows, I will advocate a greater role for experimentation.

Why are experiments the method of choice in every "hard" scientific discipline?
- They permit us to isolate and manipulate causal stimuli.
- Thereby, they permit us to make unequivocal causal inferences.

Of course, experimental control imposes tradeoffs; most notably, reduced realism and questionable generalizability. In political science, these criticisms have carried sufficient weight to retard the development of experimental research. But how significant are these criticisms today? I will argue that modern field techniques and technological advances associated with the growth of the internet go a long way toward neutralizing the traditional weaknesses of experimentation. Online experiments may prove just as realistic and generalizable as conventional sample surveys.

The Issue of Causal Inference

The founding fathers of the field of political communication (Paul Lazarsfeld, Bernard Berelson, and their successors at the University of Michigan) were all trained in survey research and accepted the logic of treating survey respondents' self-reported exposure to campaign communication as a surrogate for actual exposure. In this tradition, the standard test of a communication effect is the magnitude of differences in some criterion variable (typically vote choice) between respondents who self report high or low levels of exposure to a message. By this criterion, the effects of political campaigns were found to be negligible and "minimal consequences" became the operative wisdom among scholars studying campaigns (for a review of campaign effects research see Iyengar and Simon, 2000).

What campaign researchers failed to acknowledge, however, was the extent to which minimalist conceptions of campaign effects were inextricably bound up with their reliance on survey methods. As early as 1959, the weaknesses of survey design for isolating communication-related causes of political attitudes had been well documented. In a widely cited paper, Hovland (1959) demonstrated that surveys necessarily underestimate the effects of mass communication. Hovland identified several artifacts of survey research including unreliable and biased measures of media exposure.
The assumption that self-reported media exposure converges with actual exposure is problematic on several grounds. People have notoriously weak memories, especially when the "target" event in question concerns their encounters with political campaigns (see, for instance, Bradburn, Rips and Shevell, 1987; Pierce and Lovrich, 1982). In the Ansolabehere/Iyengar experiments, for example, over fifty percent of the participants who were exposed to a political advertisement were unable, some thirty minutes later, to recall having seen the advertisement (Ansolabehere and Iyengar, 1999). In other contexts, errors in the opposite direction may be just as likely. Thus Gilliam and I have found that because viewers of television news are so familiar with the "crime script," a script which associates violent crime with the behavior of nonwhite males, they mistakenly recall the presence of a nonwhite perpetrator even when the news report in question provides no information about the perpetrator (Gilliam and Iyengar, 2000). Naturally, the considerable error in self reports necessarily attenuates survey-based estimates of the effects of political campaigns (see Bartels, 1993).

In addition to the problem of measurement error, self-reported exposure to campaign messages is typically endogenous to political attitudes, including candidate preference. That is, those who choose to tune in to the campaign differ systematically (in ways that matter to their vote choice) from those who do not. To take the case of campaign advertising, people who can and cannot recall an advertisement are likely to differ in innumerable ways. In addition to having better memories, the former are likely to be more interested in politics, more attached to the candidates, more informed about the issues, and more likely to vote. The presence of feedback between vote intention and recall of advertising seriously undermines claims about the impact of exposure to advertising on participation. Respondents who recalled having seen a negative campaign advertisement in the 1992 National Election Study survey, for instance, were more likely to intend to vote (Wattenberg and Brians, 1999) than those who did not, leading the researchers to suggest that negative campaigning stimulated turnout. But was it exposure to negative advertising that prompted turnout, or was the greater interest in campaigns among likely voters responsible for their higher level of recall? When recall of advertising in the same survey was treated as endogenous to vote intention and the effects re-estimated using appropriate two-stage methods, the sign of the coefficient for recall was reversed: those who recalled negative advertisements were less likely to intend to vote (see Ansolabehere et al., 1999). Unfortunately, most survey-based analyses fail to disentangle the reciprocal effects of self-reported exposure to the campaign and partisan attitudes/behaviors.

In contrast to surveys, experiments provide more precise measures of media exposure that are uncontaminated by problems of endogeneity. Because exposure is manipulated and occurs prior to elicitation of the dependent measures (typically vote choice or candidate preference), and because exposure is not influenced by self-selection (random assignment), the effects of exposure can be treated as exogenous.

Precise measurement of media exposure is a prerequisite to valid causal inference. However, an even more significant advantage of experiments for the study of campaigns -- and the focus of this discussion -- is more elementary still; experiments enable researchers to isolate and test the specific components of campaigns.

At the aggregate level, campaigns encompass a concatenation of messages, channels, and sources, all of which may influence the audience, often in inconsistent directions. The researcher’s task is to identify specific causal factors and delineate the range of their relevant attributes. Even at the relatively narrow level of campaign advertisements, for instance, there are any number of potential causal factors, both verbal and visual. What was it about the infamous "Revolving Door" advertisement that is thought to have
moved so many voters in 1988? Was it, as widely alleged, Mr. Horton's race? Or was it the violent and brutal nature of his behavior, the fact that he was a convict, the race of his victim, or what? Experiments make it possible to isolate the explanation, whether it be text-based, or in the form of audio-visual cues. Returning to the case of the Horton advertisement, we could estimate the effects of race by constructing identical facial composites of Mr. Horton that vary his skin color. Alternatively, we might test for the effects of facial hair by presenting the convict's face with and without a beard. In effect, the experimentalist can decompose a message into core components which become the basis for experimental manipulations. In the next section, I describe manipulations from previous and current research designed to isolate the effects of the tone of advertising campaigns, the gender of the candidate sponsoring particular advertisements, the race of the criminal suspect in television news reports and, finally, the labeling of candidates in campaign mail.

Negativity and Candidate Reputation in Campaign Advertising

Experimental tests of negativity require variation in the tone of a campaign advertisement while holding all other attributes constant. (Although the contrast with survey-based measures of exposure to negative advertising could not be more stark, the inherent imprecision of surveys has not discouraged survey analyses of campaign negativity. For recent examples, see Geer and Finkel, 1998; Freedman and Goldstein, 1999; Kahn and Kenney, 2000.) In the Ansolabehere/Iyengar experiments, the researchers manipulated negativity by varying the text (the soundtrack) of an advertisement while preserving the visual backdrop. The negative version of the message typically placed the sponsoring candidate on the unpopular side of a salient policy issue. Thus, during the 1990 gubernatorial campaign between Pete Wilson and Dianne Feinstein, our ads described the candidates either as opponents or proponents of off-shore oil drilling and thus as either friends or enemies of the environment. This manipulation was implemented by substituting the word "yes" for the word "no." In the positive conditions, the script began as follows: "When federal bureaucrats asked for permission to drill for oil off the coast of California, Pete Wilson/Dianne Feinstein said no . . ." In the negative conditions, we substituted " . . . Pete Wilson/Dianne Feinstein said yes . . . " An additional substitution was written into the end of the ad when the announcer claimed that the sponsoring candidate would work to either "preserve" or "destroy" California’s natural beauty. A similar logic was applied to the issue of crime. The experimental ads described the candidates as having either presided over increases or decreases in violent crime. Given the consensual nature of these issues, negativity could be attributed to candidates who claimed their opponent was soft on pollution and crime.¹

Candidate: Feinstein, Issue: Environment

Tone: Positive
In addition to the tone of advertising, we varied the candidate sponsoring the ad. The theoretical basis for this manipulation was that sponsors with particular attributes would be deemed more credible on particular messages. In effect, viewers might consider one class of candidates as superior on particular policy issues. In what would prove to be a farsighted manipulation, during the 1992 campaign we compared Democrats Bill Clinton, Barbara Boxer, and Dianne Feinstein as opponents of sexual harassment in the workplace and advocates for gender equality in general. By holding the sponsor’s party constant, we were able to assess the role of gender as an antecedent of credibility. In this particular case, we embedded a sound bite from each of the sponsoring candidates. Because we were not able to match
their statements (beyond a focus on women’s rights), this particular design added realism (the appearance and voice of the sponsor) at the cost of precision. By adding the candidate’s own words to the manipulation, we substantially increased the number of confounded variables. That is, a particular word or phrase, or some physical attribute of the candidate could account for any observed differences across conditions.

**Issue: Sexual Harassment**

![Video image]

**Sponsor: Boxer**

![Video image]

**Sponsor: Feinstein**

![Video image]

**Sponsor: Clinton**

**Racial Cues in Local News Coverage of Crime**

As any regular viewer of television will attest to, crime is a frequent occurrence in the news. In response
to market pressures, television stations have adopted a formulaic approach to covering crime, an approach designed to attract and maintain the highest degree of audience interest. This "crime script" suggests that crime is invariably violent and those who perpetrate crime are disproportionately nonwhite. Because the crime script is encountered so frequently (several times each day in many cities) in the course of watching local news, it has attained the status of common knowledge. Just as we know full well what happens when one walks into a restaurant, we also know -- or at least think we know -- what happens when crime occurs.

In a series of recent experiments, Frank Gilliam and I have examined the effects of both elements of the crime script on audience attitudes. For illustrative purposes, I focus here on the racial element. In essence, we manipulated the racial appearance of the alleged suspect depicted in a typical news report while maintaining all other visual characteristics. The original stimulus was a local news report which included a close-up "mug shot" of the suspect. The picture was digitized, "painted" to alter the perpetrator’s skin color, and then re-edited into the news report. Beginning with two different perpetrators (a white male and a black male), we were able to produce altered versions of each individual in which their race was reversed, but all other features remained identical. Thus, any differences in the responses of the subjects exposed to the white or black perpetrators can only be attributed to the perpetrator’s race.

**Issue: Crime Coverage in Local News**

---

**Suspects: Black**

---

**Suspects: White**

---
My final example derives from research into candidates’ use of a particular form of direct mail known in California as "slate mail." A slate generally refers to a list of candidates nominated by a political party. Slate mail is produced and distributed by private campaign consultants, some of whom mail millions of pieces per election. The front side of a typical slate mailer includes small photographs of some of the candidates appearing on the slate, and short endorsements of particular candidates or propositions. There may also be information on the date of the election and the location of the voter's polling place, as well as other relevant logos or symbols. On the reverse side is the slate itself, namely, a listing of all the endorsed candidates and yes or no recommendations on ballot measures.

Prior to 1996, California law required candidates who paid for their names to be included in slate mail to be identified by an asterisk placed after their names. Following the passage of Proposition 208 in 1996, "paying" candidates and ballot measures were to be identified by three dollar signs following their names. On the grounds that the labeling provisions of Proposition 208 represented a significant disincentive for candidates and ballot measures to rely on slate mail, various slate mail consultants filed a constitutional challenge. After a trial, U.S. District Judge Lawrence Karlton declared certain provisions of Proposition...
208 unconstitutional and preliminarily enjoined enforcement of the entire proposition without reaching a decision on the merits of many of the challenges, including the challenges to the slate mail provisions. The Ninth Circuit Court of Appeals affirmed the preliminary injunction but returned the case to Judge Karlton for him to rule on the remaining issues. As of this writing, no further proceedings have been initiated in Judge Karlton’s court.

Following the passage of Proposition 208 Dan Lowenstein and I carried out two separate studies designed to assess the effects of dollar signs as labels for candidates. With the help of two slate mail consultants, we distributed two identical versions of Democratic and Republican slate mail in which the candidates and ballot measures that had paid to be included were identified with either the pre or post-Proposition 208 rules. (As might be expected, we found that voters tended to dislike candidates and causes associated with dollar signs.)

**Issue: Slate Mail Labelling**

![Slate Mail Cover](image1)

![Slate Mail Asterisks Label](image2)
As the examples described above demonstrate, the experiment provides unequivocal causal evidence because the researcher is able to isolate the causal factor in question, manipulate its presence or absence, and hold all other potential causes constant.

**The Issue of Generalizability**

Of course, experiments are not without their limitations. Most are administered upon "captive" populations -- college students who must serve as guinea pigs in order to gain course credit. College sophomores may be a convenient subject population, but they are not comparable to "real people." Experiments also require a somewhat sterile, laboratory-like environment which bears little resemblance to the "blooming, buzzing confusion" of the real world. In fact, the development of field (as opposed to laboratory) experiments was in direct response to the issue of mundane realism. As I discuss below, both these limitations can be overcome in the current research environment.

**Increasing Mundane Realism**
Because of the need for tightly controlled stimuli, the setting in which the typical laboratory experiment occurs is often quite dissimilar from the setting in which subjects ordinarily experience the "target" phenomenon. Concern over the realism of laboratory experiments has given rise to an increased use of so-called field experiments in which the procedures and settings more closely reflect ordinary life.

A common strategy in field experiments is the reliance on interventions with which subjects are familiar. The Ansolabehere/Iyengar campaign experiments were relatively realistic in the sense that they occurred during ongoing campaigns characterized by heavy levels of advertising (see Ansolabehere and Iyengar, 1996). The presence of a political advertisement in the local news (the vehicle used to convey the manipulation) was hardly unusual or unexpected since candidates advertise most heavily during news programs. The advertisements featured real candidates -- Democrats and Republicans, liberals and conservatives, males and females, incumbents and challengers -- as the sponsors. The material that made up the experimental stimuli were selected either from actual advertisements used by the candidates during the campaign, or were produced to emulate typical campaign advertisements. In the case of the latter, the researchers spliced together footage from actual advertisements or news reports making the treatment ads representative of the campaign advertising genre. (Of course, the need for control made it necessary for the test ads to differ from actual political ads in several important attributes including the absence of music and the appearance of the sponsoring candidate.)

Realism also depends upon the physical setting in which the experiment is administered. Asking subjects to report to a location on a university campus may suit the researcher, but may make the experience of watching television equivalent to the experience of visiting the doctor. A more realistic strategy is to provide subjects with a milieu that closely matches the setting of their living room or den. To that end, the Ansolabehere/Iyengar experimental "laboratory" was designed to resemble, as closely as possible, the natural "habitat" of the television viewer. Comfortable couches and chairs were arranged in front of a television set, with houseplants and wall hangings placed around the room. Respondents had access to refreshments and reading matter (newspapers and magazines) during the viewing sessions. In most cases, a family member or friend took part in the experiment at the same time, so that subjects did not find themselves sitting next to a stranger while viewing the political advertisements.4

A further step toward realism concerns the power of the manipulation (also referred to experimental realism). Naturally, the researcher would like the manipulation to be noticed by the subject. At the same time, it is important that the manipulation not overpower the subject (as in the Milgram obedience studies where the task of administering electric shock to a fellow subject proved overpowering to many). In the case of the campaign advertising experiments, we resolved the power of the manipulation -- mundane realism tradeoff by embedding the manipulation in a commercial break of a local newscast. The political ad appeared with other non-political ads (we excluded other political ads from the newscast) and because subjects were led to believe that the study was about "selective perception of news," we gave them no rationale to pay particular attention to ads. Overall, the manipulation was relatively small, amounting to thirty seconds of a fifteen minute videotape.

In general, the mundane realism - experimental control trade off is significant. The fact that we encouraged subjects to watch the treatments with someone they knew meant that their level of familiarity with fellow subjects is subject to unknown variation. And, as in the example of the ads dealing with sexual harassment, producing experimental ads that more closely emulated actual ads introduced a series of confounded variables associated with the appearance and voice of the sponsoring candidate.

Reducing Sampling Bias
In conventional experimental research, it is possible to recruit non-student participants, but at considerable cost/effort. Locating experimental facilities at public locations and enticing a quasi-representative sample to participate proves both cost- and labor-intensive. Typical costs include rental fees for an experimental facility in a public area (such as a shopping mall) where it is possible to attract a wide range of participants; recruitment and compensation of subjects; and recruitment, training and compensation of research staff to administer the experiments. In the local news experiments conducted in Los Angeles in 1998 and 1999, the total costs per subject amounted to approximately $45. Fortunately, technology has both enlarged the pool of potential participants and reduced the per capita cost of administering subjects.

Today, traditional experimental methods can be rigorously and far more efficiently replicated using online strategies. The advantages of using the internet as the experimental "site" include the ability to reach diverse populations without geographic limitations. The rapid development of multimedia-friendly web browsers make it possible to bring video presentations to the computer screen. Indeed, all the major presidential candidates now "air" their televised ads at their web sites, and it is common practice for people to follow news reports from online sources such as CNN. The technology is so accessible that subjects can easily "self-administer" experimental manipulations. Compared with conventional shopping mall studies, therefore, the costs are minimal. Moreover, with the ever-increasing use of the internet not only are the samples more diverse, the process by which participants encounter the manipulation (logging on and surfing the Web) is also more realistic.

In current research at Stanford University, we have been examining the demographic profiles of experimental "samples" recruited online. To date, our data are limited to "drop-in" subjects -- subjects who managed to navigate themselves to the Political Communication Laboratory web site and who then signed up to participate in a survey or experiment. We are in the process of broadening the online participant pool by using banner ads that offer potential subjects a financial incentive to participate. Naturally, we anticipate that subjects who respond to a monetary incentive will prove more diverse than those who volunteer.

The demographic make-up of our online subjects is shown in Graph 1. As a baseline, we provide the same indicators for a representative sample of adult Americans with home access to the internet as well as a representative sample of all voting-age adults.5 (The survey was administered in March 2000.) The first comparison enables us to gauge the degree of online self-selection; that is, the degree of divergence between our participants and typical internet users. The second comparison indicates the degree of discrepancy between our self-selected online samples and a representative sample of voting-age adults.

Graph 1

Comparison of Online Experimental Participants, National Sample of Americans with Internet Access, and National Sample of Voting-Age Adults
The comparisons in Graph 1 support two broad patterns. The more surprising of the two is that the participants in the online experiments reasonably approximate the online user population. The digital divide, however, is still sufficiently wide to exclude major categories of the population from participating in online studies.

The comparison between our experimental subjects and the representative sample of Americans with home internet access was closest with respect to race/ethnicity and education. The predominance of whites and the college-educated was about the same in the subject pool and among national survey respondents with internet access. (In fact, a slightly higher percent of the national sample scored "post-college" on education.)

Of the background factors considered, political communication researchers are most interested in the variable of party identification. Here again, experimental subjects proved representative of the broader user population in that independents and non-partisans were the largest group, followed by Republicans and Democrats. Although the relative share of the three groups was consistent in the experiments and survey, the experimental pool was somewhat less Democratic (and more Republican) than the broader online population. Notwithstanding these minor differences, it appears that party identification does not enter into the decision to participate in media online experiments.

The clearest evidence of experimental selection bias emerged with age and gender. Study participants were much younger (on average, by ten years), and much more likely to be male than the online population. The sharp divergence in age may be attributed to the fact that our studies are launched from on an academic server that is more likely to be encountered by college students, and also to the general
"surfing" proclivities of younger users. The gender gap is more puzzling and may reflect differences in political interest. Our studies are explicitly political in focus, which may act as a disincentive to potential women subjects.

In summary, if the population of interest consists of Americans with online access, the experimental participants comprise a reasonably representative sample at least with respect to race, education, and party affiliation. The experiments deviate from the online population on the attributes of gender and age, drawing disproportionately male and younger participants.

Turning to the digital divide, the access threshold remains a strong liability for online research. In relation to the broader adult population, our experimental participants are younger, more educated, more likely to be white males, and less apt to identify as a Democrat. With the exception of age and gender, these differences are just as stark when the comparison is between all adults and the representative sample of adults with online access (for evidence of the scale of differences between internet users and nonusers, see Moss and Mitra, 1999; Papadakis, 2000).

Demographic differences associated with internet access are well known. The question of parallel attitudinal differences, however, has received less attention. Our data suggest that online access is correlated with political party identification. Study participants (and online users in general) are less likely to identify with either party and, among those who do identify, Republicans enjoy a clear advantage over Democrats. The plurality of Republicans online may reflect access differences associated with economic standing and ethnicity. Whatever the reason, online researchers interested in the effects of campaigns will need to make special efforts to oversample Democrats.

Although these data make it clear that people who participate in online media experiments are no microcosm of the adult population, the fundamental advantage of online over conventional field experiments cannot be overlooked. Conventional experiments recruit subjects from particular locales, online experiments draw subjects from across the country. The Ansolabehere/Iyengar campaign experiments, for example, recruited subjects from a particular area of southern California (greater Los Angeles). The online experiments, in contrast, attracted subjects from thirty different American states and several countries.

**Conclusion**

The standard comparison of experiments and surveys favors the former on the grounds of precise causal inference and the latter on the grounds of greater generalizability. As I have suggested, however, traditional experimental methods can be effectively, and just as rigorously replicated using online strategies which have the advantage of reaching a participant pool that is more far-flung and diverse than the pool relied on by conventional experimentalists. The data shown above tend to underestimate this advantage because the studies in question were based on the most restrictive of selection procedures (pure self selection). It is possible to reduce the dispositional basis of study participation by altering the mix of incentives. Using gift coupons or cash credit vouchers as inducements for participation can only serve to boost the diversity of the participant pool.

Online techniques also permit a more precise "targeting" of recruitment procedures so as to enhance participant diversity. Banner ads publicizing the study and the financial incentives for study participants can be placed in portals or sites that are known to attract underrepresented groups. Women subjects or African Americans, for instance, could be attracted by ads placed in sites catering to these groups.
The most compelling argument in favor of online experiments, however, is the inexorable diffusion of information technology. As online communication penetrates classrooms across America and as the market share of online communication sources grows, the external validity gap between experimental and survey methods can only close.

Endnotes

1 Of course, this approach assumes a one-sided distribution of policy preferences and the tone manipulation would be reversed for participants in the experiment who actually favored off shore drilling and increased crime rates.


3 California Prolife Council Political Action Committee v. Scully, 164 F-3d 1189 (9th Cir.1999).

4 It is possible, of course, for the experimental setting to be too realistic. During the early days of our campaign experiments, we provided subjects with access to a remote control device, only to discover that a subject used it to fast forward the tape during the commercial breaks.

5 I am grateful to Doug Rivers, CEO of Inter-Survey, for making these data available.

6 Overseas participants have been excluded from the data presented in Graph 1.

References


