

Experimental Designs for Political Communication Research: Using New Technology and Online
Participant Pools to Overcome the Problem of Generalizability

Shanto Iyengar
Stanford University
Political Communication Lab
<http://pcl.stanford.edu>

To appear in E. P. Bucy and R. L. Holbert eds., *Sourcebook for Political Communication Research: Methods, Measures, and Analytical Techniques*. Routledge, forthcoming.

Introduction

As recently as three decades ago, the use of experimental methods was a rarity in the disciplines of political science, sociology, and communications. Beginning in the early 1980s, a surge of interest in the interdisciplinary field of political psychology set in motion a trickle of experimental methods into several sub-fields of political science, including political communication. But despite the increased interest, longstanding concerns over the artificiality of experimental settings, the unrepresentativeness of experimental subject pools and the questionable generalizability of experimental findings has continued to impede the further diffusion of experimental methods.

In this chapter I describe the inherent strengths of the experiment as a basis for causal inference, using recent examples from political communication research. I argue that the technological advances associated with the rapid diffusion of the Internet have already gone a long way toward neutralizing the traditional weaknesses of experimentation. First, experiments administered online can prove just as realistic as conventional experiments. Second, issues of sampling bias -- previously endemic to experiments -- can be overcome through the greater “reach” of online experiments and, in addition, by the application of standard probability sampling techniques to the recruitment of online experimental participants. These developments significantly alleviate concerns over the generalizability of experimental research and as a result, experiments now represent a dominant methodology for political communication researchers.

Causal Inference: The Strength of Experiments

In the field of political communication, the principal advantage of the experiment over the survey -- and the focus of the discussion that follows -- is the researcher's ability to isolate and test the effects of specific components of political messages. Consider the case of political 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 virtually an infinite number of potential causal factors, both verbal and visual. What was it about the infamous "Willie Horton" advertisement that is thought to have moved so many American voters away from Michael Dukakis during the 1988 presidential campaign? Was it, as widely alleged, Mr. Horton's race? Or was it the violent and brutal nature of his described behavior, the fact that he was a convict, the race of his victim, or something else entirely? Experiments make it possible to isolate the explanation, whether it be text-based or in the form of audio-visual cues. Surveys, on the other hand, can only provide indirect evidence on self-reported exposure to the causal variable in question.

Of course, experiments not only shed light on treatment effects but also enable researchers to test more elaborate hypotheses concerning the interaction of message factors with individual difference variables. Not all individuals are equally susceptible to incoming messages. Perhaps Democrats with a weak party affiliation and strong sense of racial prejudice were especially likely to sour on Governor Dukakis in the aftermath of exposure to the Horton advertisement.

The weaknesses of survey design for isolating the effects of mass communication have been amply documented. In a widely cited paper, Hovland (1959) identified several problematic artifacts of survey research including unreliable measures of media exposure. Clearly, exposure is a necessary pre-condition for media influence, but self-reported exposure to media coverage is hardly equivalent to actual exposure. People have notoriously weak memories for political experiences (see, for instance, Bradburn, Rips and Shevell, 1987; Pierce and Lovrich, 1982). In the Ansolabehere and Iyengar experiments on campaign advertising (which spanned the 1990, 1992, and 1994 election cycles), over 50% of the participants who were exposed to a political advertisement were unable, *some thirty minutes later*, to recall having seen the advertisement (Ansolabehere, 2006). In a more recent example, Vavreck found that nearly half of a control group not shown a public service message responded either that they couldn't remember or that they *had* seen it (Vavreck, 2007; also see, Prior, 2003). Errors of memory also compromise recall-based measures of exposure to particular news stories (see Gunther, 1987) or news sources (see Price and Zaller, 1993). Of course, the scale of the error in self-reports necessarily attenuates survey-based estimates of the effects of political campaigns (see Bartels, 1993, 1996; Prior, 2003).

An even more serious obstacle to causal inference in the survey context is that self-reported media exposure is typically endogenous to a host of political attitudes researchers seek to explain, including candidate preference. That is, those who claim to read newspapers or watch television news on a regular basis differ systematically (in ways that matter to their vote choice) from those who do not. This problem has become more acute in the aftermath of the revolution in information technology. In 1968,

approximately 75% of the adult viewing audience watched one of the three network evening newscasts, but by 2008 the combined audience for network news was less than 35% of the viewing audience. In 2008, the only people watching the news were those with a keen interest in politics; most everyone else had migrated to more entertaining, non-political programming alternatives (see Prior, 2007).

The endogeneity issue has multiple ramifications for political communication research. First, consider those instances where self-reported exposure is correlated with political predispositions but actual exposure is not. This is generally the case with televised political advertising. Most voters encounter political ads unintentionally, in the course of watching their preferred television programs in which the commercial breaks contain a heavy dose of political messages. Thus, actual exposure is idiosyncratic (based on the viewer's preference for particular programs), while self-reported exposure is based on political predispositions.

The divergence in the antecedents of self-reported exposure has predictable consequences for "effects" research. In experiments that manipulated the tone of campaign advertising, Ansolabehere and Iyengar (1995) found that actual exposure to negative messages "demobilized" voters, i.e., discouraged turnout. However, on the basis of self-reports, survey researchers concluded that exposure to negative campaign advertising stimulated turnout (Wattenberg and Briens, 1999). But was it recalled exposure to negative advertising that prompted turnout, or 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 express an intention 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 and behaviors. As this example suggests, in cases where actual exposure is less selective than reported exposure, self-reports may prove especially biased.

In other scenarios, the tables may be turned and the experimental researcher may be at a disadvantage. Actual exposure to political messages in the real world is typically not analogous to random assignment. Unlike advertisements, news coverage of political events can be avoided by choice, meaning that exposure is limited to the politically engaged strata. Thus, as Hovland (1959) pointed out, manipulative control actually weakens the ability to generalize to the real world where exposure to politics is typically voluntary. In these cases, as I note later, it is important that the researcher use designs that combine manipulation with self-selected exposure.

In summary, the fundamental advantage of the experimental approach is the ability to isolate causal variables, which remain the basis for experimental manipulations. In the next section, I describe manipulations designed to isolate the effects of negative advertising campaigns, racial cues in television news coverage of crime, and the physical similarity of candidates to voters.

Negativity in Campaign Advertising

At the very least, establishing the effects of negativity in campaign advertising on voters' attitudes requires varying the tone of a campaign advertisement while holding all other attributes of the advertisement constant. Despite the significant increase in

scholarly attention to negative advertising, few if any studies live up to this minimal threshold of control (for representative examples of survey-based analyses see Geer and Finkel, 1998; Freedman and Goldstein, 1999; Kahn and Kenney, 2000.)

In a series of experiments conducted by Ansolabehere and Iyengar, the researchers manipulated negativity by varying the text (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 some salient policy issue. Thus, during the 1990 gubernatorial campaign between Pete Wilson (Republican) and Dianne Feinstein (Democrat), the treatment ads positioned the candidates either as opponents or proponents of offshore oil drilling and thus as either friends or foes of the environment. This manipulation was implemented by simply 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 “said yes” for “said no.” An additional substitution was written into the end of the ad when the announcer stated that the candidate in question would either work to “preserve” or “destroy” California’s natural beauty. Given the consensual nature of the issue, negativity could be attributed to candidates who claimed their opponent was soft on polluters.¹

The results from these studies (which featured gubernatorial, mayoral, senatorial, and presidential candidates) indicated that participants exposed to negative rather than positive advertisements were less likely to intend to vote. The demobilizing effects of exposure to negative advertising were especially prominent among viewers who did not identify with either of the two political parties (see Ansolabehere and Iyengar, 1995).

Racial Cues in Local News Coverage of Crime

As any regular viewer of television will attest to, crime is a frequent occurrence in broadcast 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 (Gilliam and Iyengar, 2000).

In a series of recent experiments, researchers have documented the effects of both elements of the crime script on audience attitudes (see Gilliam, Valentino and Beckman, 2002; Gilliam, Iyengar, Simon, and Wright, 1996). For illustrative purposes, I focus here on the racial element. In essence, these studies were designed to manipulate the race/ethnicity of the principal suspect depicted in a news report while maintaining all other visual characteristics. The original stimulus consisted of a typical local news report, which included a close-up still “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. As shown below, beginning with two different perpetrators (a white male and a black male), the researchers were able to produce altered versions of each individual in which their race was reversed, but all other features remained identical. Participants who watched the news report in which the suspect was thought to be non-white expressed greater support for “punitive” policies (e.g., imposition of “three strikes and you’re out”

remedies, treatment of juveniles as adults, and support for the death penalty). Given the precision of the design, these differences in the responses of the subjects exposed to the white or black perpetrators could only be attributed to the perpetrator's race (see Gilliam and Iyengar, 2000).

Figure 1: Race of Suspect Manipulation



Facial Similarity as a Political Cue

A consistent finding in the political science literature is that voters gravitate to candidates who most resemble them on questions of political ideology, issue positions, and party affiliation. But what about physical resemblance; are voters also attracted to candidates who look like them?

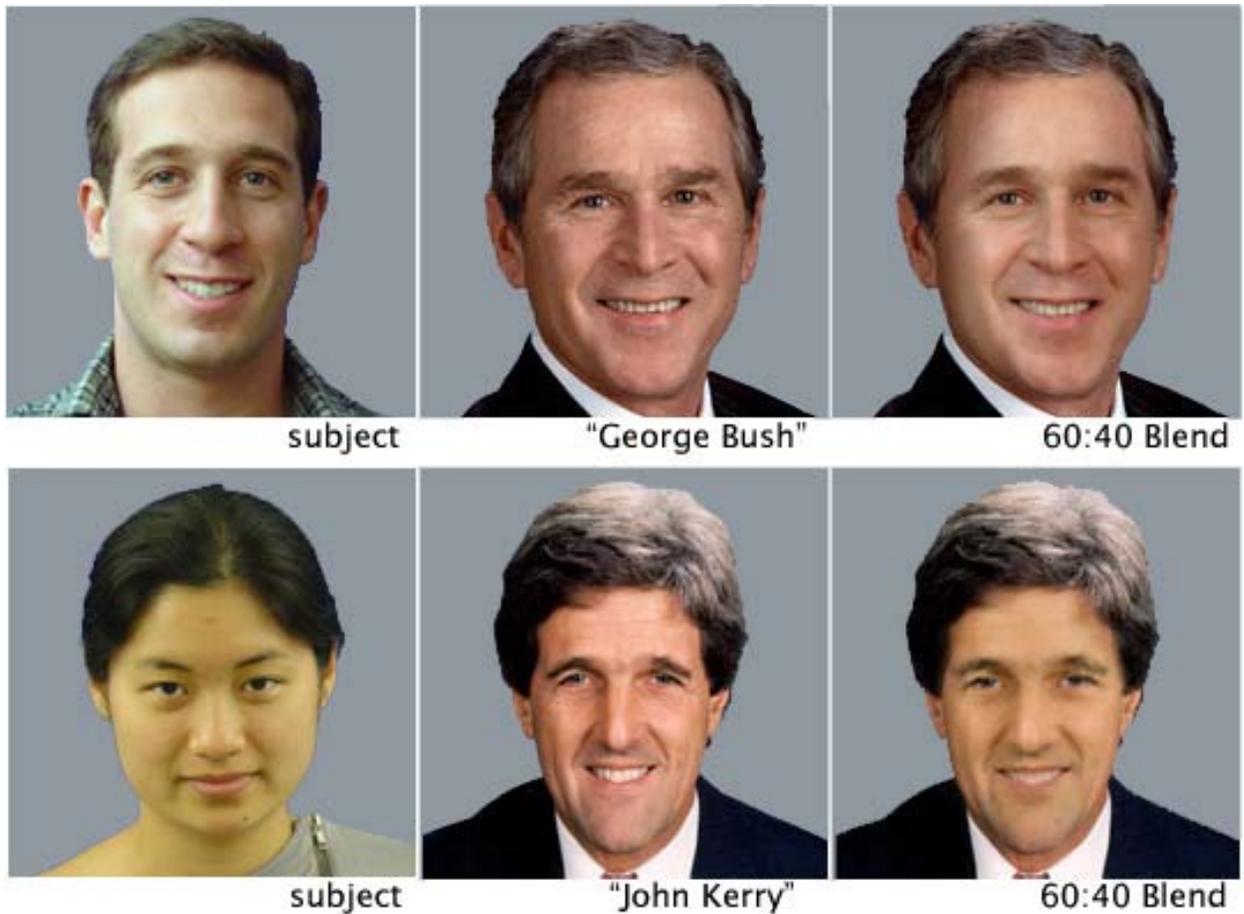
Several lines of research suggest that physical similarity in general, and facial

similarity in particular, is a relevant criterion for choosing between candidates. Thus, frequency of exposure to any stimulus – including faces – induces a preference for that stimulus over other, less familiar stimuli (Zajonc, 2001). Moreover, evolutionary psychologists argue that physical similarity is a kinship cue and there is considerable evidence that humans are motivated to treat their kin preferentially (see, for instance, Burnstein, Crandall, and Kitayama, 1994; Nelson, 2001).

In order to isolate the effects of facial similarity on voting preferences, researchers obtained digital photographs of 172 registered voters selected at random from a national Internet panel (for details on the methodology, see Bailenson, Iyengar, and Yee, 2008). Participants were asked to provide their photographs approximately three weeks in advance of the 2004 presidential election. One week before the election, these same participants were asked to participate in an online survey of political attitudes that included a variety of questions about the presidential candidates (President George W. Bush and Senator John Kerry). The screens for these candidate questions included photographs of the two candidates displayed side by side. Within this split-panel presentation, participants had their own face either morphed with Bush or Kerry at a ratio of 60% of the candidate and 40% of themselves.² Figure 2 shows two of the morphs used in this study.

Figure 2

The Facial Similarity Manipulation



The results of the face morphing study revealed a significant interaction between facial similarity and strength of the participant's party affiliation. Among strong partisans, the similarity manipulation had no effect; these voters were already convinced of their vote choice. But weak partisans and independents -- whose voting preferences were not as entrenched -- moved in the direction of the more similar candidate (see Bailenson, Iyengar, and Yee, 2008). Thus, the evidence suggests that non-verbal cues can influence voting, even in the most visible and contested of political campaigns.³

In short, as these examples indicate, 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 other potential causes constant. Any observed differences between experimental and control groups, therefore, can only be attributed to the factor that was manipulated.

Improved Measures of Audience Response

The ability to launch experiments online further strengthens the ability of communication researchers to draw causal inferences by providing more precisely calibrated indicators of audience reactions to media messages. In contrast with conventional experiments, which typically provide only verbal measures of audience response, online experiments permit observation of information seeking behavior and enable finely-grained, longitudinal indicators of voter response to campaign advertisements.

Behavioral Indicators of Selective Exposure

Researchers have long assumed that people possess an innate preference for attitude-consistent messages or sources of information. According to this “selective exposure” hypothesis, voters seek to avoid information that clashes with their preexisting beliefs (e.g., Festinger 1957) and instead put themselves in the path of information they expect to agree with. As Lazarsfeld et al., (1948) pointed out, biased exposure to information has clear implications for the exercise of informed citizenship: “In recent years there has been a good deal of talk by men of good will about the desirability and necessity of guaranteeing the free exchange of ideas in the market place of public opinion. Such talk has centered upon the problem of keeping free the channels of expression and communication. Now we find that the consumers of ideas, if they have

made a decision on the issue, themselves erect high tariff walls against alien notions” (89).

Given the practical difficulties of delivering large quantities of information, the typical study on selective exposure provides participants with only a limited range of choice. As Cotton pointed out, the literature has failed to address “how people actively seek and avoid information on their own” (1985, 29) in naturalistic settings. Digital technology now makes it possible to deliver voluminous quantities of information in a compact and easy to navigate format.

In a study of selective exposure during the 2000 presidential campaign, researchers provided a representative sample of registered voters with a multimedia CD containing extensive information about candidates Bush and Gore – including text of all of their stump speeches delivered between July 1 and October 7, a full set of televised ads, and the texts of the Democratic and Republican party platforms. The CD also included the soundtrack and transcripts of the candidates’ nomination acceptance speeches and the first televised debate. All told, the information amounted to over 600 pages of text and two hours of multimedia (see Iyengar et al., 2008).

The campaign CD was delivered to a representative sample of American adult Internet users two weeks before Election Day. Participants were informed in advance that their use of the CD would be examined by the researchers (and they were requested not to share the CD with members of their family or friends). As the user navigated through the CD offerings, a built-in tracking feature recorded every visited page (in the order of visit), the number of total times the CD was accessed, and the length of each browsing session in a log file on the user’s hard drive. Upon completing a post-election

questionnaire, participants were given instructions for finding and uploading their log-files. From these files, we were able to monitor the degree to which CD users gravitated to information provided by the candidate they preferred. The findings revealed only partial evidence of selective exposure based on partisanship; Republicans (and conservatives) showed a lop-sided preference for information concerning Bush, but Democrats (and liberals) proved more even-handed in their information-seeking behavior.

The tendency for partisans on the right to show greater avoidance of attitude-discrepant information is attributable to both dispositional and contextual factors. In comparison with liberals, conservatives may have a more intense sense of group identity, thus heightening their need to avoid dissonance. On the other hand, the greater selectivity among Republicans may reflect habituation over time. Since the launch of the Fox Network in 1986, Republicans have enjoyed easy access to news with a pro-Republican tilt. This experience with Fox News may have encouraged similar information-seeking behavior in a non-news context.

Continuous Tracking of Viewers' Reactions to Campaign Ads

Campaign advertising is the major source of information for voters in non-presidential elections. Understanding voters' reactions to ads is thus fundamental to understanding the effectiveness of campaigns. Most researchers who investigate the effectiveness of ad campaigns typically rely on verbal measures to gauge the influence of ads. Viewers might be asked if they agreed or disagreed with the ad in question, or if the ad elicited positive or negative feelings concerning the sponsoring candidate. These

measures ask respondents to provide a summary or “averaged” assessment of their reaction to the content and imagery of ads.

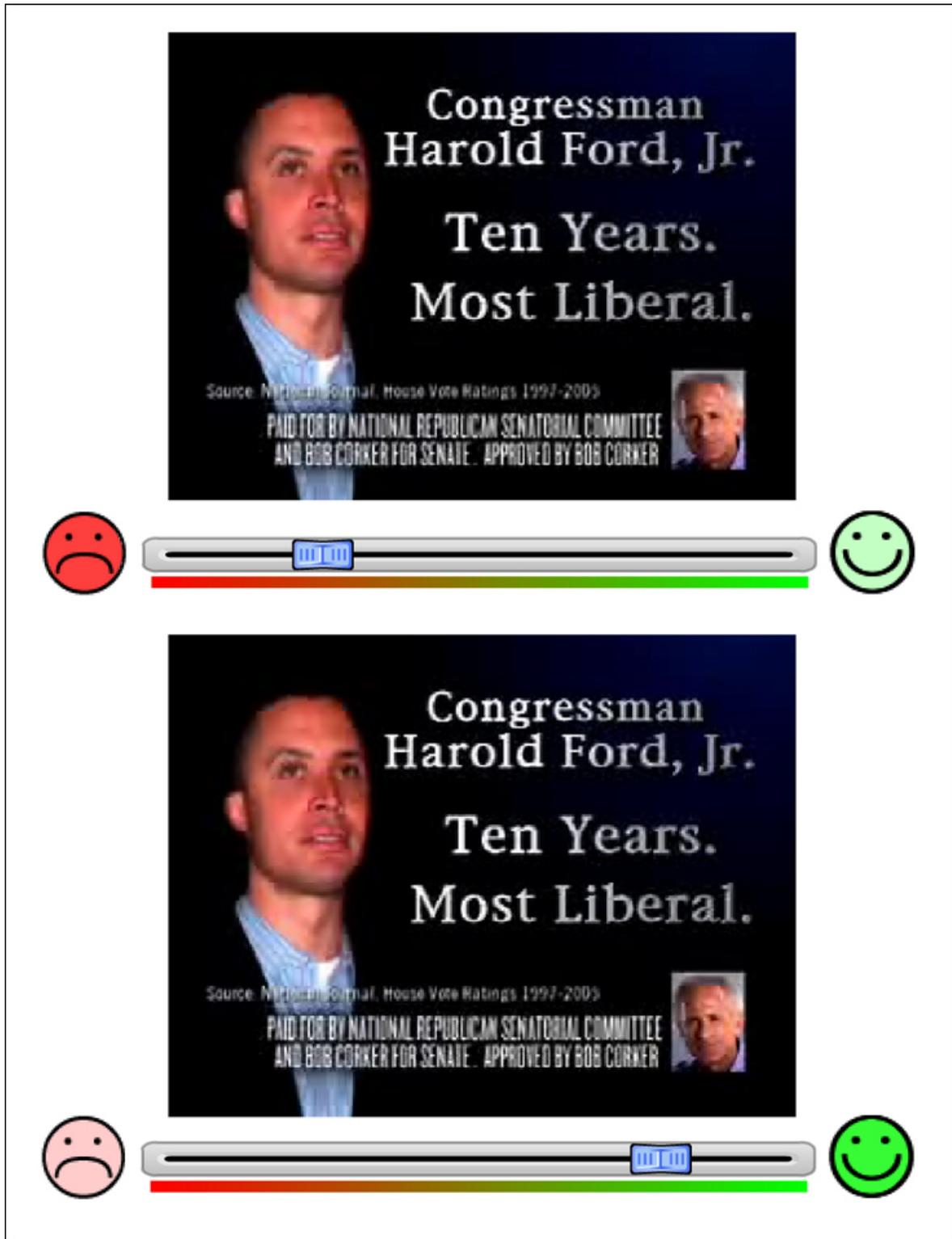
With the diffusion of digital technology, it is possible to monitor viewer response to advertising on a continuous basis, over the entire playing of the ad (see Iyengar, Jackman, and Hahn, 2007). Rather than asking for a summary assessment *after* the participant has watched the ad, researchers can use an online “dialing” procedure that synchronizes viewers’ self-reported feelings concerning the soundtrack and visual imagery they encounter at any given moment *during* the playing of the ad.

The dial meter methodology was implemented online in a study of the 2006 senate elections in six battleground states. A sample of approximately 1900 registered voters with Internet access was selected at random from a nationwide online panel. Participants were instructed (and given a practice task) on how to move a slider located immediately below the video in accordance with their feelings about the content of the ad. The specific instruction was: If what you see or hear makes you feel good, or you agree with the speaker, indicate this by moving the slider towards the green end. If, however, your reaction is negative, and you dislike what you see or hear, then move the slider to the red zone.”

Special Javascript-based software recorded the position of the slider once a second by evenly dividing the range of dial positions into 100 intervals, with zero indicating the left or negative end of the dial, and 100 the right or positive end. Thus, as the ad played, we could monitor voters’ reactions in real time from beginning to end. At the start of each ad, the slider was positioned at the neutral or “50” position, and this was the first dial value recorded for each ad view. Figure 3 displays a pair of screen shots

from one of the Tennessee conditions featuring the race between Republican Bob Corker and Democrat Harold Ford Jr., with two hypothetical settings of the dial.

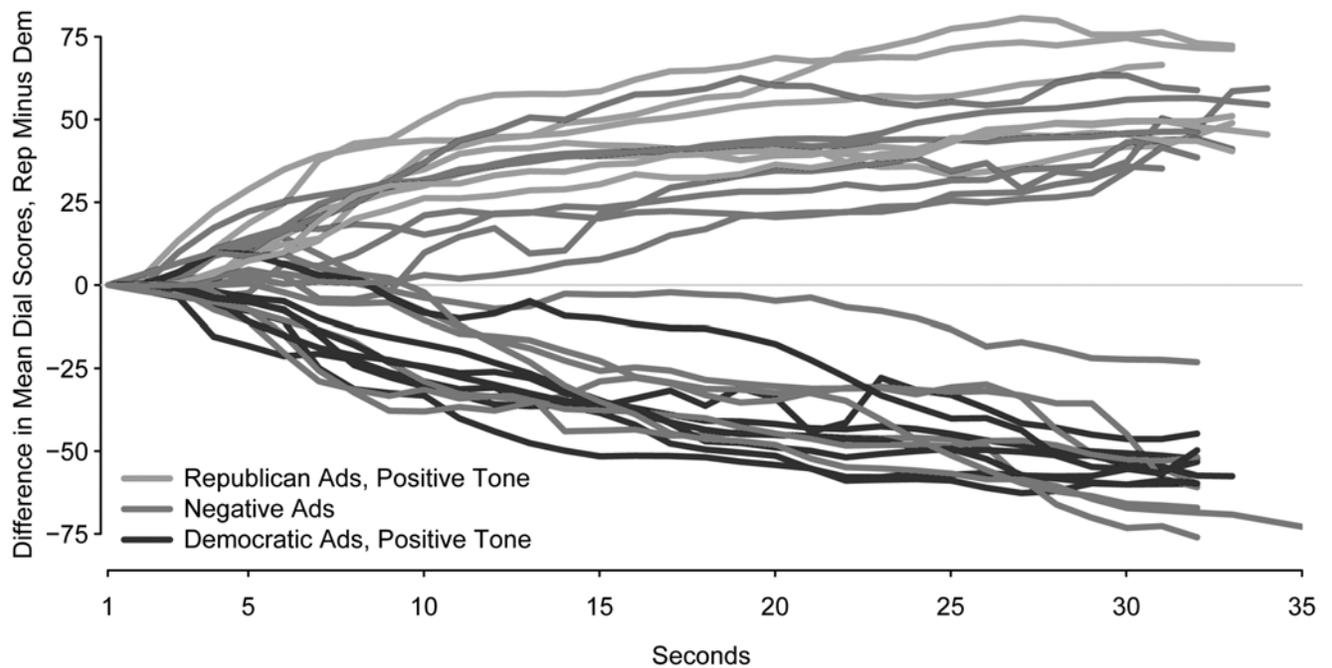
Figure 3: Screen Shots from Online Dials



As the ad played, participants could move the slider to indicate their feelings about the content of the ad, with the position of the dial recorded once a second.

The results from this study indicated that most ads polarize partisan viewers; over the course of the ad, Democrats and Republicans inevitably move in opposite directions (see Figure 4). This pattern is consistent with prior research which shows that exposure to campaign ads strengthens viewers' partisan predispositions (Ansolabehere and Iyengar, 1995). While partisans responded rapidly to the content of advertising, independents were typically unmoved, remaining ambivalent over the entire playing of the ad.

Figure 4: Partisan Polarization in Dial Scores



A further finding from this dials study was that the rate of polarization proved variable across the partisanship of the sponsoring candidate. Democrats consistently converged (arrived at their stable end point) faster in response to Democratic ads than did Republicans in response to Republican ads. In effect, Democratic ads resonated more

powerfully with Democrats than Republican ads did with Republicans. Perhaps this effect was due to the partisan appeal of the ads' messages. Democratic ads, which tended to highlight the state of the war in Iraq, the fallout from the Abramoff scandal and which linked the Republican candidate with President Bush, mobilized the Democratic base more effectively than generic Republican appeals on national security, immigration, or taxes.

The Issue of Generalizability

The problem of limited generalizability, long the bane of experimental design, is manifested at three levels: the realism of the experimental setting, the representativeness of the participant pool, and the discrepancy between experimental control and self-selected exposure to media presentations.

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 artificial properties 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.⁴

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 televised advertising (see Ansolabehere and Iyengar, 1995). 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 genre. (The need for control made it necessary for our stimulus 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 also 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.

A further step toward realism concerns the power of the manipulation (also referred to as experimental realism). Of course, the researcher would like for the manipulation to have an effect. At the same time, it is important that the required task or stimulus not overwhelm the subject (as in the Milgram obedience studies where the task of administering an electric shock to a fellow participant proved overpowering and ethically suspect). In the case of the campaign advertising experiments, we resolved the experimental realism versus mundane realism tradeoff by embedding the manipulation in a commercial break of a local newscast. For each treatment condition, the stimulus ad appeared with other non-political ads and because subjects were led to believe that the study was about “selective perception of news,” they had no incentive to pay particular attention to ads. Overall, the manipulation was relatively small, amounting to thirty seconds of a fifteen-minute videotape.

In general, there is a significant tradeoff between experimental realism and manipulative control. In the advertising studies described above, the fact that subjects were exposed to the treatments in the company of others meant that their level of familiarity with fellow subjects was subject to unknown variation. And producing experimental ads that more closely emulated actual ads (e.g. ads with musical background included and featuring the sponsoring candidate) would necessarily have introduced a series of confounded variables associated with the appearance and voice of the sponsor. Despite these tradeoffs, however, it is still possible to achieve a high degree of experimental control with stimuli that closely resemble the naturally occurring phenomenon of interest.

Sampling Bias

The most widely cited limitation of experiments concerns the composition of the subject pool (Sears, 1986). Typically, laboratory experiments 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 for academic researchers, but are they comparable to "real people?"

In conventional experimental research, it is possible to broaden the participant pool 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), recruitment of participants, and training and compensation of research staff to administer the experiments. In our local news experiments conducted in Los Angeles in the summer and fall of 1999, the total costs per subject amounted to approximately \$45. Fortunately, as described below, technology has both enlarged the pool of potential participants and reduced the per capita cost of administering an experimental study.

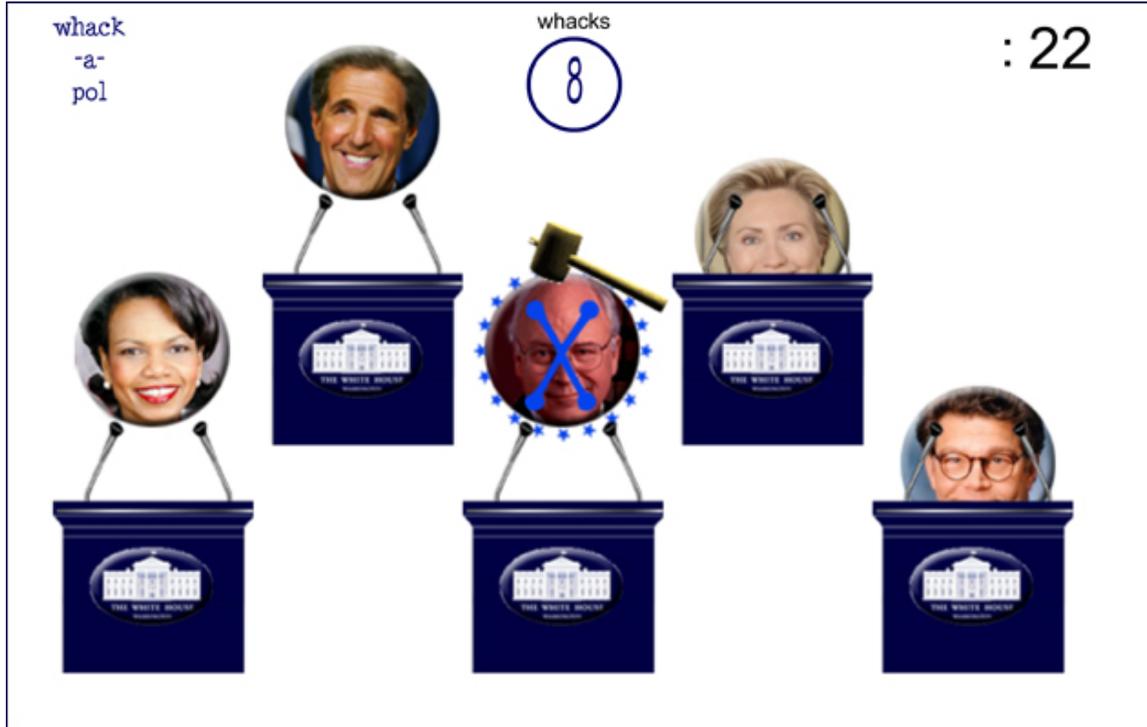
Today, traditional experimental methods can be rigorously and far more efficiently administered using an online platform. Utilizing the Internet as the experimental "site" provides several advantages over conventional locales including the ability to reach diverse populations without geographic limitations. Advances in multimedia have made it possible to bring video presentations to the computer screen with relative ease. Indeed, it is now standard for candidates to feature their televised ads on their web sites, and access to video reports from online sources is growing. The

technology is sufficiently user friendly that most Web users can easily “self-administer” experimental manipulations. Compared with conventional shopping mall studies, therefore, the costs of online experiments are minimal. Moreover, with the ever-increasing use of the Internet, not only are the samples more diverse but the setting in which participants encounter the manipulation (surfing the Web on their own) is also more realistic.

“Drop-in” Samples

The Political Communication Laboratory at Stanford University has been administering experiments over the Internet for nearly a decade. One of the Lab’s more popular online experiments is “whack-a-pol” (<http://pcl.stanford.edu/exp/whack/polm>), modeled on the well-known whack-a-mole arcade game. Ostensibly, the game provides participants with the opportunity to “bash” well-known political figures. Before playing the game, participants complete a consent form and brief pretest survey. After playing the game, they self-administer the post-test survey. Since the game imposes severe time and attention constraints (players see five different moving faces, each hittable for a period of between two and three seconds), the whacking task provides an unobtrusive measure of group identity. That is, we expect subjects to target “out-group” figures (defined on the basis of party identification, i.e., Democrats should target Republicans and vice-versa) for more extensive whacking. Party affiliation is the most salient basis for political choice, although one could replicate this design with other groupings such as gender or ethnicity.

Figure 5: Whack-a-Pol Screenshot



Since going live in 2001, over 2000 visitors have played whack-a-pol. These “drop in” subjects found the PCL site on their own initiative. How does this group compare with a representative sample of adult Americans with home access to the Internet, and a representative sample of all voting-age adults? We can use two different baselines to assess this. First, we gauge the degree of online self-selection; that is, the degree of divergence between drop-in participants and typical Internet users. The second comparison indicates the degree of discrepancy between self-selected online samples and all voting-age adults (from a sample drawn in 2000).

The results of these comparisons showed two broad patterns. The more surprising of the two was that the participants in the online experiments reasonably approximated the online user population. However, as described below, there is still evidence of a significant digital divide in the sense that major categories of the population remain

underrepresented in online studies.

The match between our drop-in experimental subjects and the typical Internet user was closest with respect to race/ethnicity and education. The predominance (over 80%) of whites and the college-educated (50%) in the online participant pool was approximately the same as among all Internet users. The key background variable of party identification also proved consistent across online participants and the user community at large. Among whack-a-pol players, Republicans were the largest group (37%), followed by Democrats and Independents. Although the participant pool was somewhat less Democratic (and more Republican) than the broader online population, it appears that party identification is not a significant predictor of the decision to participate in online experiments with political content.

The clearest evidence of selection bias vis-à-vis the online population emerged with age and gender. The mean age of study participants was 31 (it was 41 for the online sample) and participants were also more likely to be male (65%) than the typical online user (52%). The sharp divergence in age may be attributed to the fact that our studies are launched from 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 female subjects, who are known to be less interested in politics than men.

In summary, if the population of interest is limited to American Internet users, the results of the online experiments can be generalized at least with respect to race, education, and party affiliation. Experimental participants deviate from the online

population on the attributes of gender and age, drawing disproportionately male and younger participants.

Turning to the comparisons between online participants and the population at large (the digital divide question), it is clear that the technology access threshold remains a strong liability for online research. In relation to the broader adult population, our experimental participants were 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 were just as stark when the comparison was between the offline and online populations (for evidence of the scale of differences between Internet users and nonusers, see Papadakis, 2000; Pew Internet and American Life Project, 2005).

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 advertising experiments, for example, recruited subjects from a particular area of southern California (greater Los Angeles). The online experiments, in contrast, attracted a sample of subjects from thirty different American states and several countries.

Expanding the Pool of Online Participants

One way to broaden the online subject pool is by recruiting participants from more well-known and frequently visited websites. News sites that cater to political junkies, for example, may be motivated to increase their “circulation” by collaborating with scholars whose research studies focus on controversial issues. While the researcher

obtains data which may be used for scholarly purposes, the website gains a form of “interactivity” through which the audience may be engaged. Playing an arcade game or watching a brief video clip may pique participants’ interest thus encouraging them to return to the site and boosting the news organization’s online traffic.

In recent years, PCL has partnered with Washingtonpost.com to expand the reach of online experiments. Studies designed by PCL -- focusing on topics of interest to people who read Washingtonpost.com -- are advertised on the Website’s “politics” section (see Figure 6). Readers who click on a link advertising the study in question are sent directly to the PCL site, where they complete the experiment, and are then returned to Washingtonpost.com. The results from these experiments were then described in a newspaper story and online column. In cases where the results were especially topical (e.g., a study of news preferences described below showing that Republicans avoided CNN and NPR in favor of Fox News), a correspondent from Washingtonpost.com hosted an online “chat” session to discuss the results and answer questions.

Figure 6: Link Advertising Washingtonpost.com Experiment

SOCIAL SCIENCES SURVEY



Post-Stanford Experiments

What Do You Think of Political Ads?
With the 2006 midterm elections approaching, washingtonpost.com and Stanford University want to get your views on new campaign ads.
Watch the Ads and Take the Survey.

The development of cable television and explosion of media outlets on the Internet have created a more fragmented information environment in which cable news, talk radio, and 24-hour news outlets compete for attention. Given this dramatic increase in the number of available news outlets, it is not surprising that media choices increasingly reflect partisan considerations. In the new media environment, there is growing evidence that partisans gravitate to sources perceived as more congenial to their preferences (see Pfau et al., 2007).

A PCL-Washingtonpost.com experiment was designed to investigate the relationship between political predispositions and news preferences. This study assessed the extent to which partisans on the right treated Fox News as a preferred provider. More specifically, the study assessed whether attention to the identical news story was increased or decreased when the story was attributed to Fox News, NPR, CNN or the

BBC.

Using the MSNBC daily news feed (which includes news reports from a variety of sources), news reports were randomly assigned to one of four sources -- Fox, NPR, CNN, or BBC. Participants were provided a brief headline accompanied by the logo of the news organization and asked to indicate which of the four reports displayed on the screen they would like to read. (They could also click a "can't say" box, as shown in Figure 7.) They repeated this task across six different news categories -- American politics, the war in Iraq, "race in America," crime, travel, and sports. We also included a baseline or control condition in which all source logos were deleted; here participants could only choose between the reports based on the text of the headlines. All other aspects of the presentation were equalized across the different news organizations. For instance, the placement of a particular story or source on the screen was randomized so that no particular news source gained from being the first or last on the screen.

Figure 7: Screenshot from News Preferences Experiment

Which of the following articles about POLITICS are you most interested in reading?

<input type="radio"/>		Rumsfeld heckled during speech Protesters repeatedly interrupted Defense Secretary Donald H. Rumsfeld during a speech Thursday and a former CIA analyst accused him of lying about Iraq prewar intelligence in an unusual display of anti-war sentiment.
<input type="radio"/>		Senate OKs \$109 billion war, Katrina bill The Senate on Thursday passed a \$109 billion bill to pay for the war in Iraq and hurricane aid, but a presidential veto threat imperils many provisions added by lawmakers.
<input type="radio"/>		CIA Director Goss resigns CIA Director Porter Goss is resigning, President Bush said Friday.
<input type="radio"/>		Republican right abandoning Bush A new Associated Press poll shows that angry conservatives are driving the approval ratings of President Bush and the GOP-led Congress to dismal new lows.
<input type="radio"/>	Can't Say	

The results from this experiment revealed that the effects of the source manipulation on news story selection were strongest for political topics where partisan divisions are intense. Thus, Republicans gravitated to Fox reports on politics and Iraq, while Democrats avoided Fox in favor of CNN or NPR. Even though the partisan divide was bigger for hard news, it did not disappear entirely for non-political subjects. Republicans continued to show a preference for Fox News, even when reading about the NBA or possible vacation destinations (for details, see Iyengar and Hahn 2007).

To date, each of the Washingtonpost.com – PCL joint experiments have succeeded in attracting relative large samples, at least by the standards of experimental research.⁶ Experiments on especially controversial or newsworthy subjects attracted a high volume of traffic (on some days exceeding 500). In other cases, the rate of

participation slowed to a trickle, resulting in a longer period of time to gather the data.

Sampling from Online Research Panels

Even though drop-in online samples provide more diversity than the typical “college sophomore” sample, they are obviously biased in several important respects. Participants from Washingtonpost.com, for instance, included very few conservatives or Republicans. Fortunately, it is now possible to overcome issues of sampling bias -- assuming the researcher has access to funding -- by administering online experiments to representative samples. In this sense, the lack of generalizability associated with experimental designs is largely overcome.

Two market research firms have pioneered the use of Web-based experiments with fully representative samples. Not surprisingly, both firms are located in the heart of Silicon Valley. The first is Knowledge Networks based in Menlo Park, and the second is Polimetrix (recently purchased by the UK polling company of YouGov) based in Palo Alto.

Knowledge Networks has overcome the problem of selection bias inherent to online surveys (which reach only that proportion of the population that is both online and inclined to participate in research studies) by recruiting a nationwide panel through standard telephone methods. This representative panel (including over 150,000 Americans between the ages of 16 and 85) is provided free access to the Internet via a WebTV. In exchange, panel members agree to participate (on a regular basis) in research studies being conducted by Knowledge Networks. The surveys are administered over the panelist’s WebTV. Thus, in theory Knowledge Networks can deliver samples that meet the highest standards of probabilistic sampling. In practice, because their panelists have

an obligation to participate, Knowledge Networks also provides relatively high response rates (Dennis et al., 2004).

Polimetrix uses a novel “matching” approach to the sampling problem. In essence, they extract a quasi-representative sample from large panels of online volunteers. The process works as follows. First, Polimetrix assembles a very large pool of opt-in participants by offering small incentives for study participation (e.g. the chance of winning an Ipod). As of November, 2007 the number of Polimetrix panelists exceeded 1.5 million Americans.

In order to extract a representative sample from this pool of self-selected panelists, Polimetrix uses a two-step sampling procedure. First, they draw a conventional random sample from the target population of interest (i.e. registered voters). Second, for each member of the target sample, Polimetrix substitutes a member of the opt-in panel who is similar to the corresponding member of the target sample on a set of demographic characteristics such as gender, age, and education. In this sense, the matched sample consists of respondents who “represent” the respondents in the target sample. Rivers (2006) describes the conditions under which the matched sample approximates a true random sample.

The Polimetrix samples have achieved impressive rates of predictive validity, thus bolstering the claims that matched samples emulate random samples. In the 2005 California special election, Polimetrix accurately predicted the public’s acceptance or rejection of all seven propositions (a record matched by only one other conventional polling organization) with an average error rate comparable to what would be expected given random sampling (Rivers, n.d.).

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. 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. Female subjects or African Americans, for instance, could be attracted by ads placed in sites tailored to their interests. Most recently, the development of online research panels make it possible to administer experiments on broad cross-sections of the American population. As technology diffuses still further, the generalizability gap between experimental and survey methods will continue to close.

Although technology has clearly advanced the conduct of experimental research, there are challenges ahead. The most notable concerns the increasingly self-selected nature of media audiences. Since there is a much wider range of media choices than ever before, providing greater variability in the content of available information, people uninterested in politics can avoid news programming altogether while political junkies can exercise discretionary or selective exposure to political information.

The self-selected composition of audiences has important consequences for political communication research. In a recent study, for example, young voters in

California were mailed a CD featuring the candidates contesting the 2002 gubernatorial election (see Iyengar and Jackman, 2003). The researchers found that actual turnout among young voters who used the CD was eleven percentage points higher than among young voters in the control group (who did not receive the CD). This observed difference could be attributed not only to the treatment, but also to the ex-ante level of political interest among participants who chose to use the CD. When exposure to the experimental treatment is based on choice, it becomes necessary to estimate the average treatment effect after adjusting for self-selection. In the CD experiment, 78 percent of those assigned to the treatment group ignored the CD, due to general disinterest in the subject matter, insufficient time, or other such factors. Those who did accept the treatment were drawn disproportionately from the ranks of those interested in politics. Thus, exposure to the treatment was non-random and correlated with key outcome variables of interest.

Fortunately, in recent years there has been considerable progress in estimating treatment effects in non-randomized experimental or observational settings. Recent reviews include Imbens (2004), Angrist and Krueger (2000), and Heckman, Ichimura and Todd (1998). The general idea is straightforward: although respondents have self-selected into treatment, after the researcher controls for factors that predispose assignees to accept or refuse treatment, the outcomes of interest and treatment are no longer confounded. Given the availability of variables (covariates) known to motivate participation, the researcher can overcome the failure of random assignment and recover an unbiased estimate of the treatment effect. In particular, it is possible to carry out *matched comparisons* of treated and control participants (matching on the covariates);

averaging over these matched comparisons generally produces an unbiased estimate of the causal effect of treatment (see Rosenbaum and Rubin, 1983).

In the youth CD study, the researchers were able to match the treatment and control groups for previous voting history, propensity to participate in surveys, and socio-economic indicators related to political participation (i.e., age, marital status, and education). In comparison with non-participants, CD users were older, more frequent survey takers, more educated, and had higher incomes. After adjusting for these motivational biases, the effect of CD use on actual turnout was five percent, a substantial reduction from the original estimate of eleven percent.

In sum the use of digital technology in experimental research is a double-edged sword. While researchers are in a position to administer precisely controlled manipulations to an increasingly diverse subject pool, thus overcoming the traditional weakness of experiments, they face a radically altered media environment in which exposure to political content is driven by choice (see Bennett and Iyengar, 2008). As a result, assignment to treatments in political communication-related experiments will inevitably depend on the participant's political preferences, and estimating treatment effects will require the use of more powerful statistical tools.

References

- Angrist, Joshua and Alan Krueger. "Empirical Strategies in Labor Economics," pp. 1277-1366 in *Handbook of Labor Economics: Vol 3*, eds. Orley Ashenfelter and David Card. New York: Elsevier Science, 2000.
- Ansolabehere, Stephen. "The paradox of minimal effects," pp. 29-44 in *Capturing Campaign Effects*, eds. Henry Brady and Richard Johnston. Ann Arbor: University of Michigan Press, 2006.
- Ansolabehere, Stephen, and Shanto Iyengar. *Going Negative: How Political Ads Shrink and Polarize the Electorate*. New York: Free Press, 1995.
- Ansolabehere, Stephen, Shanto Iyengar, and Adam Simon. "Replicating Experiments Using Aggregate and Survey Data." *American Political Science Review*, 93 (December 1999): 901-10.
- Bailenson, Jeremy, Iyengar, Shanto, Yee, Nick, and Nathan Collins. "Facial Similarity Between Candidates and Voters Causes Influence." *Public Opinion Quarterly*, in press.
- Larry Bartels. "Messages Received: The Political Impact of Media Exposure." *American Political Science Review* 87 (June 1993): 267-85.
- Bennett, W. Lance, and Shanto Iyengar. "A New Era of Minimal Effects? The Changing Foundations of Political Communication." *Journal of Communication*, in press.
- Bradburn, Norman, Lance Rips, and Steven Shevell. "Answering Autobiographical Questions: The Impact of Memory and Inference in Surveys." *Science* 236 (April 1987): 157-61.

- Burnstein, Eugene, Christine Crandall, and Shinobu Kitayama. "Some Neo-Darwinian Decision Rules for Altruism: Weighing Cues for Inclusive Fitness as a Function of the Biological Importance of the Decision." *Journal of Personality and Social Psychology* 67 (November, 1994): 773-789.
- Cotton, John. "Cognitive Dissonance in Selective Exposure," pp. 11-33 in *Selective Exposure to Communication*, eds. Dolf Zillman and Jennings Bryant. Hillsdale, NJ: Lawrence Erlbaum, 1985.
- Dennis, J. Michael, Rick Li, and Cindy Chatt. 2004. Benchmarking Knowledge Networks' web-enabled panel survey of selected GSS questions against GSS in-person interviews. Knowledge Networks Technical Report, February.
- Festinger, Leon. *A Theory of Cognitive Dissonance*. New York: John Wiley, 1957.
- Finkel, Steven, and John Geer. "A Spot Check: Casting Doubt on the Demobilizing Effect of Attack Advertising." *American Journal of Political Science* 42 (April 1998): 573-95.
- Freedman, Paul, and Kenneth Goldstein. "Measuring Media Exposure and the Effects of Negative Campaign Ads." *American Journal of Political Science* 43 (October 1999): 1189-208.
- Gerber, Alan, and Donald Green. "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment." *American Political Science Review* 94 (September 2000): 653-63.
- Gilliam, Franklin Jr., and Shanto Iyengar. "Prime Suspects: The Influence of Local Television News on the Viewing Public." *American Journal of Political Science* 44 (July 2000): 560-73.

Gilliam, Franklin Jr., Nicholas Valentino, and Mathew Beckman. "Where You Live and What You Watch: The Impact of Racial Proximity and Local Television News on Attitudes About Race and Crime." *Political Research Quarterly* 55 (December 2002): 755-80.

Gilliam, Franklin Jr, Shanto Iyengar, Adam Simon, and Oliver Wright. 1996. "Crime in Black and White: The Violent, Scary World of Local News." *Harvard International Journal of Press/Politics* 1 (July 1996): 6-23.

Gunther, Barrie. 1987. *Poor Reception: Misunderstanding and Forgetting Broadcast News*. Hillsdale, NJ: Lawrence Erlbaum.

Heckman, James, Hidehiko Ichimura and Petra Todd. "Matching as an Econometric Evaluation Estimator." *Review of Economic Studies* 65 (April 1998): 261-94.

Hovland, Carl. "Reconciling Conflicting Results Derived From Experimental and Survey Studies of Attitude Change." *American Psychologist* 14 (January 1959): 8-17.

Imbens, Guido. "Semiparametric Estimation of Average Treatment Effects under Exogeneity: A Review." *Review of Economics and Statistics* 86 (February, 2004): 4-29.

Iyengar, Shanto, and Kyu Hahn. "Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use." Paper Presented at the Annual Meeting of the International Communication Association, May, 2007.

Iyengar, Shanto, and Simon Jackman. "Can Information Technology Energize Voters? Evidence from the 2000 and 2002 Campaigns." Paper Presented at the Annual Meeting of the American Political Science Association, August, 2003.

Iyengar, Shanto, and Adam Simon. "New Perspectives and Evidence on Political

- Communication and Campaign Effects,” pp. 149-69 in *Annual Review of Psychology*, 51. Palo Alto, CA: Annual Reviews Press, 2000.
- Iyengar, Shanto, Kyu Hahn, Jon Krosnick, and John Walker. “Selective Exposure to Campaign Communication: The Role of Anticipated Agreement and Issue Public Membership.” *Journal of Politics* 70 (January 2008): 186-200.
- Kahn, Kim, and Patrick Kenney. “Do Negative Campaigns Mobilize or Suppress Turnout? Clarifying the Relationship between Negativity and Participation.” *American Political Science Review* 93 (December 1999): 877-90.
- Lazarsfeld, Paul, Bernard Berelson, and Hazel Gaudet. *The People’s Choice*. New York: Columbia University Press, 1948.
- Nelson, Charles. The development and neural bases of face recognition: Reply to Critiques. *Infant and Child Development* 10 (April 2001): 3-18.
- Papadakis, Maria. “Complex Picture of Computer Use in the Home Emerges.” National Science Foundation Issue Brief, March 31, 2000.
- Pew Internet and American Life Project. “Digital Divisions,” Available at http://www.pewinternet.org/pdfs/PIP_Digital_Divisions_Oct_5_2005.pdf
- ____. “Home Broadband Adoption,” Available at http://www.pewinternet.org/pdfs/PIP_Broadband_trends2006.pdf
- ____. “Parent and Teenager Internet Use,” Available at http://www.pewinternet.org/pdfs/PIP_Teen_Parents_data_memo_Oct2007.pdf
- Pfau, Michael, Brian Houston, and Shane Semmler. *Mediating the Vote: The Changing Media Landscape in U.S. Presidential Campaigns*. Lanham, MD: Rowman and Littlefield, 2007.

- Pierce, John, and Nicholas Lovrich. "Survey Measurement of Political Participation: Selective Effects of Recall in Petition Signing." *Social Science Quarterly* 63 (March 1982): 164-71.
- Price, Vincent, and John Zaller. "Who Gets the News? Alternative Measures of News Reception and Their Implications for Research." *Public Opinion Quarterly* 57 (Summer 1993): 133-64.
- Prior, Markus. "Any Good News in Soft News? The Impact of Soft News Preference on Political Knowledge." *Political Communication* 20 (April-June 2003): 149-72.
- _____. *Post-Broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Elections*. New York: Cambridge University Press, 2007.
- Douglas Rivers. Sample Matching: Representative Sampling from Internet Panels.
http://www.polimetrix.com/documents/Polimetrix_Whitepaper_Sample_Matching.pdf.
- _____. Scientific Sampling for Online Research.
http://www.polimetrix.com/documents/Polimetrix_Sampling.pdf
- Rosenbaum, Paul, and Donald Rubin. "The Central role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70 (January 1983): 41-55.
- David Sears. "College Sophomores in the Laboratory: Influences of a Narrow Data Base on the Social Psychology View of Human Nature." *Journal of Personality and Social Psychology* 51 (March 1986): 515-30.
- Snyder, Mark, John Grather, and Kristine Keller. "Staring and Compliance: A Field Experiment on Hitchhiking." *Journal of Applied Social Psychology* 4 (June

1974): 165-70.

Vavreck, Lynn. "The Exaggerated Effects of Advertising on Turnout: The Dangers of Self-Reports." *Quarterly Journal of Political Science* 2 (June 2007): 325-43.

Wattenberg, Martin, and Craig Brians. "Negative Campaign Advertising: Demobilizer or Mobilizer?" *American Political Science Review* 93 (December 1999): 891-900.

Zajonc, R. "Mere Exposure: A Gateway to the Subliminal." *Current Directions in Psychological Science*, 10 (December 2001): 224-228.

Endnotes

1. Of course, this approach assumes a one-sided distribution of policy preferences and that the tone manipulation would be reversed for experimental participants who actually favored off shore drilling.
2. We settled on the 60:40 ratio after a pretest study indicated that this level of blending was insufficient for participants to detect traces of themselves in the morph, but sufficient to move evaluations of the target candidate.
3. Facial similarity is necessarily confounded with familiarity – people are familiar with their own faces. There is considerable evidence (see Zajonc, 2001) that people prefer familiar to unfamiliar stimuli. An alternative interpretation of these results, accordingly, is that participants were more inclined to support the more familiar-looking candidate.
4. Psychologists typically use the term to describe experiments administered in naturalistic public settings such as elevators, subway cars or shopping malls (see Snyder et al., 1974).
5. It is possible, of course, for the experimental setting to be too realistic. During the initial implementation 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.

The number of participants per experimental session ranged from one to four. In most cases, sessions with multiple participants consisted of people who scheduled the session together (i.e. people who were members of the same family or work group). It is possible that the social atmosphere of the viewing sessions (watching with people you know, watching with strangers, or watching alone) may have interacted with the manipulations, but we had no apriori basis for expecting the effects of advertising tone on political attitudes to be conditioned by the sociability of the viewing experience.

6. Since 2006, PCL and Washingtonpost.com have undertaken eight joint studies with an average sample size of 1,300.