

Laboratory Experiments in Political Science

Shanto Iyengar
(siyengar@stanford.edu)

Departments of Political Science and Communication
Stanford University

Conference on Experimentation in Political Science, Northwestern University

August 14, 2009

Revised Version

Until the middle of the twentieth century, the discipline of political science was primarily qualitative – philosophical, descriptive, legalistic, and typically reliant on case studies that failed to probe causation in any measurable way. The word “science” was not entirely apt.

In the 1950s, the discipline was transformed by the behavioral revolution, spearheaded by advocates of a more social scientific, empirical approach. Even though experimentation was the sine qua non of research in the hard sciences and in psychology, the method remained a mere curiosity among political scientists. For behavioralists interested in individual-level political behavior, survey research was the methodology of choice on the grounds that experimentation could not be used to investigate real-world politics (for more detailed accounts of the history of experimental methods in political science, see Bositis and Steinel 1987; Kinder and Palfrey 1993; Green and Gerber 2003). The consensus view was that laboratory settings were too artificial and that experimental subjects were too unrepresentative of any meaningful target population for experimental studies to be valid. Further, many political scientists viewed experiments -- which typically necessitate the deception of research subjects -- as an inherently unethical methodology.

The bias against experimentation began to weaken in the 1970s when the emerging field of political psychology attracted a new constituency for interdisciplinary research. Laboratory experiments gradually acquired the aura of legitimacy for a small band of scholars working at the intersection of the two disciplines.¹ Most of these

¹ An important impetus to the development of political psychology was provided by the Psychology and Politics Program at Yale University. Developed by Robert Lane, the program provided formal training in psychology to political science graduate students

scholars focused on the areas of political behavior, public opinion and mass communication, but there were also experimental forays into the fields of international relations and public choice (Hermann and Herman 1967; Riker 1967). Initially, these researchers faced significant disincentives to applying experimental methods -- most importantly, research based on experiments was unlikely to see the light of day simply because there were no journals or conference venues that took this kind of work seriously.

The first major breakthrough for political scientists interested in applying the experimental method occurred with the founding of the journal *Experimental Study of Politics* in 1970. The brainchild of the late James Dyson (then at Florida State University) and Frank Scioli (then at Drew University and now at the National Science Foundation), *ESP* was founded as a boutique journal dedicated exclusively to experimental work. The co-editors and members of their editorial board were committed behavioralists who were convinced that experiments could contribute to more rigorous hypothesis testing and thereby to theory building in political science (Scioli 2009). As stated by the editors, the mission of the journal was to “provide an outlet for the publication of materials dealing with experimental research in the shortest possible time, and thus to aid in rapid dissemination of new ideas and developments in political research and theory.”

ESP served as an important, albeit specialized, outlet for political scientists interested in testing propositions about voting behavior, presidential popularity, mass communication and campaigns, or group decision making. The mere existence of a

and also hosted postdoctoral fellows interested in pursuing inter-disciplinary research. Later directors of this training program included John McConahay and Donald Kinder.

journal dedicated to experimental research (with a masthead featuring established scholars from highly ranked departments)² provided a credible signal to graduate students and junior faculty (this author included) that it might just be possible to publish (rather than perish) and build a career in political science on the basis of experimental research.

Although *ESP* provided an important “foot in the door,” the marginalized status of experiments in political science persisted during the 1970s. Observational methods, most notably, survey research, dominated experimentation even among the practitioners of political psychology. One obvious explanation for the slow growth rate in experimental research was the absence of necessary infrastructure. Experiments are typically space-, resource-, and labor-intensive. Laboratories with sophisticated equipment or technology, and trained staff were nonexistent in political science departments, with one notable exception, namely, the State University of New York at Stony Brook.

When SUNY–Stony Brook was established in the early 1960s, the political science department was given a mandate to specialize in behavioral research and experimental methods. In 1978, the department moved into a new building with state-of-the-art experimental facilities including laboratories for measuring psycho-physiological responses (modeled on the psychophysiology labs at Harvard), cognitive or information-processing labs for tracking reaction time, and an array of social psychological labs

² Scholars who played important editorial roles at *ESP* included Marilyn Dantico (who took over as co-editor of the journal when Scioli moved to NSF), Richard Brody, Gerald Wright, Heinz Eulau, James Stimson, Steven Brown and Norman Luttbeg.

modeled on the lab run by the eminent Columbia psychologist Stanley Schachter.³ Once these labs were put to use by the several prominent behavioralists who joined the Stony Brook political science faculty in the early 1970s (including Milton Lodge, Joseph Tanenhaus, Bernard Tursky and John Wahlke), the department would play a critical role in facilitating and legitimizing experimental research.⁴

The unavailability of suitable laboratory facilities was but one of several obstacles facing the early experimentalists. An equally important challenge was the recruitment of experimental subjects. Unlike the field of psychology, where researchers could draw on a virtually unlimited captive pool of student subjects, experimentalists in political science had to recruit volunteer (and typically unpaid) subjects on their own initiative. Not only did this add to the costs of conducting experiments, it also ensured that the resulting samples would be far from typical.

In the early 1980s, experimental methods were of growing interest to researchers in several subfields of the discipline. Don Kinder and I were fortunate enough to receive generous funding from the National Institutes of Health and the National Science Foundation for a series of experiments designed to assess the effects of network news on

³ The social psychology laboratories included rooms with transparent mirrors and advanced video and sound editing systems.

⁴ The extent of the Stony Brook political science department's commitment to interdisciplinary research was apparent in the department's hiring of several newly-minted social psychologists. The psychologists recruited out of graduate school -- none of whom fully understood, at least during their job interview, why a political science department would see fit to hire them -- included John Herrstein, George Quattrone, Kathleen McGraw and Victor Otatti. Of course, the psychologists were subjected to intense questioning by the political science faculty over the relevance and generalizability of their research. In one particularly memorable encounter, following a job talk on the beneficial impact of physical arousal on information processing and judgment, an expert on voting behavior asked the candidate whether he would suggest requiring voters to exercise prior to voting.

public opinion. These experiments, most of which were administered in a dilapidated building on the Yale campus, revealed that contrary to the conventional wisdom at the time, network news exerted significant effects on the viewing audience. We reported the full set of experimental results in *News That Matters* (Iyengar and Kinder 1987). The fact that the University of Chicago Press published a book based exclusively on experiments demonstrated that experiments could be harnessed to address questions of political significance. That the book was generally well received demonstrated that a reliance on experimental methodology was no longer stigmatized in political science.

By the end of the 1980s, laboratory experimentation had become sufficiently recognized as a legitimate methodology in political science for mainstream journals to regularly publish papers based on experiments (see Druckman, Green, Kuklinski, and Lupia 2006). Despite the significant diffusion of the method, however, two key concerns contributed to continued scholarly skepticism. First, experimental settings were deemed lacking in mundane realism -- the experience of participating in an experiment was sufficiently distinctive to preclude generalizing the results to real-world settings. Second, student-based and other volunteer subject pools were considered unrepresentative of any broader target population of interest (i.e. registered voters or individuals likely to engage in political protest). To this day, the problem of external validity or questionable generalizability continues to impede the adoption of experimentation in political science.

In this chapter I begin by describing the inherent strengths of the experiment as a basis for causal inference, using recent examples from my own work in political communication. I argue that the downside of experiments -- the standard “too artificial” critique -- has been weakened by several developments, including the use of more

realistic designs that move experiments outside of a laboratory environment and the technological advances associated with the Internet. The online platform is itself now entirely realistic (given the extensive daily use of the Internet by ordinary individuals); it also allows researchers to overcome the previously profound issue of sampling bias. All told, these developments have gone a long way toward alleviating concerns about the validity of experimental research -- so much so that I would argue that experiments now represent a dominant methodology for researchers in several fields of political science.

Causal Inference: The Strength of Experiments

The principal advantage of the experiment over the survey or other observational methods -- and the focus of the discussion that follows -- is the researcher's ability to isolate and test the effects of specific components of specific causal variables. 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 the potential causal mechanisms 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 forces, 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 during the campaign, that Horton was African-American (see Mendelberg 2001)? Or was it the violent and brutal nature of his described behavior, the fact that he was a convict or something else entirely? Experiments make it possible to isolate the attributes of messages that move audiences, whether these are text-based or

non-verbal 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 moderator variables by assessing interactions between the treatment factors and relevant individual-difference variables. In the case of persuasion, for instance, not all individuals are equally susceptible to incoming messages (see Zaller 1992). In the case of the 1988 campaign noted above, 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.

In contrast with the experiment, the inherent weaknesses of the survey design for isolating the effects of causal variables have been amply documented. In a widely cited paper, Hovland (1959) identified several problematic artifacts of survey-based studies of persuasion 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, Pierce and Lovrich 1982; Bradburn, Rips and Shevell 1987). 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 and Iyengar 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 (Price and Zaller 1993). Of course, the scale of the error in self-reports tends to be systematic (respondents are prone to overstate their media exposure), survey-based estimates of the effects of political campaigns are necessarily attenuated (Bartels 1993; Prior 2003).

An even more serious obstacle to causal inference in the survey context is that the indicators of the causal variable (self-reported media exposure in most political communication studies) are typically endogenous to a host of outcome variables researchers seek to explain (such as candidate preference). Those who claim to read newspapers or watch television news on a regular basis, for instance, differ systematically (in ways that matter to their vote choice) from those who attend to the media less frequently. This problem has become especially acute in the aftermath of the revolution in “new media.” 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 (Prior 2007).

The endogeneity issue has multiple ramifications for political communication research. First, consider those instances where self-reported media 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 television 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 intentions to vote. 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, Iyengar and Simon 1999).⁵

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 to the treatment is less selective than self-reported exposure, self-reports may prove especially biased.

In other scenarios, however, the tables may be turned and the experimental researcher may actually be at a disadvantage. Actual exposure to political messages in

⁵ In a meta-analysis of political advertising research, Lau et al. concluded that experimental studies were not more likely to elicit evidence of significant effects (Lau et al. 1999). The meta-analysis, however, combines experiments that utilize a variety of designs most of which fail to isolate the negativity of advertising.

the real world is typically not analogous to random assignment. People who choose to participate in experiments on campaign advertising are likely to differ from those who choose to watch ads during campaigns (for a general discussion of the issue, see Gaines and Kuklinski 2008). 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) and others (Heckman and Smith 1995) have pointed out, manipulative control actually weakens the ability to generalize to the real world where exposure to politics is typically voluntary. In these cases, it is important that the researcher use designs that combine manipulation with self-selected exposure.

One other important aspect of experimental design that contributes to strong causal inference is the provision of procedures to guard against the potential contaminating effects of “experimental demand” -- cues in the experimental setting or procedures that convey to participants what is expected of them (for the classic account of demand effects, see Orne 1962). Demand effects represent a major threat to internal validity: participants are motivated to respond to subtle cues in the experimental context suggesting what is wanted of them rather than to the experimental manipulation itself.

The standard precautions against experimental demand include disguising the true purpose of the study by providing participants with a plausible (but false) description,⁶ using relatively unobtrusive outcome measures, and maximizing the “mundane realism”

⁶ Of course, the use of deception in experimental research necessitates full “debriefing” of participants at the conclusion of the study. Typically, participants are provided with a relatively detailed account of the experiment and given the opportunity to receive any papers based on the study data. In recent years, experimental procedures have become highly regulated by university review boards in order to maximize the principle of informed consent and to preclude any lingering effects of deception. Most informed consent forms, for instance, alert participants to the use of deception in experimental research.

of the experimental setting so that participants' are likely to mimic their behavior in real-world settings. (I will return to the theme of realism later in the section on generalizability.)

In the campaign advertising experiments described below, for instance, the researchers inserted manipulated political advertisements into the ad breaks of the first ten minutes of a local newscast. Study participants were diverted from the researchers' intent by being misinformed that the study was about "selective perception of television news." The use of a design in which the participants answered the survey questions only after exposure to the treatment further guarded against the possibility that they might see through the cover story and infer the true purpose of the study.

In summary, the fundamental advantage of the experimental approach -- and the reason experimentation is the methodology of choice in the hard sciences -- is the researcher's ability to isolate causal variables, which constitute the basis for experimental manipulations. In the next section, I describe manipulations designed to assess 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 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 unobtrusively 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 California 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.⁷

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 say they intended to vote. The demobilizing effects of exposure to negative advertising were especially prominent among viewers

⁷Of 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.

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?

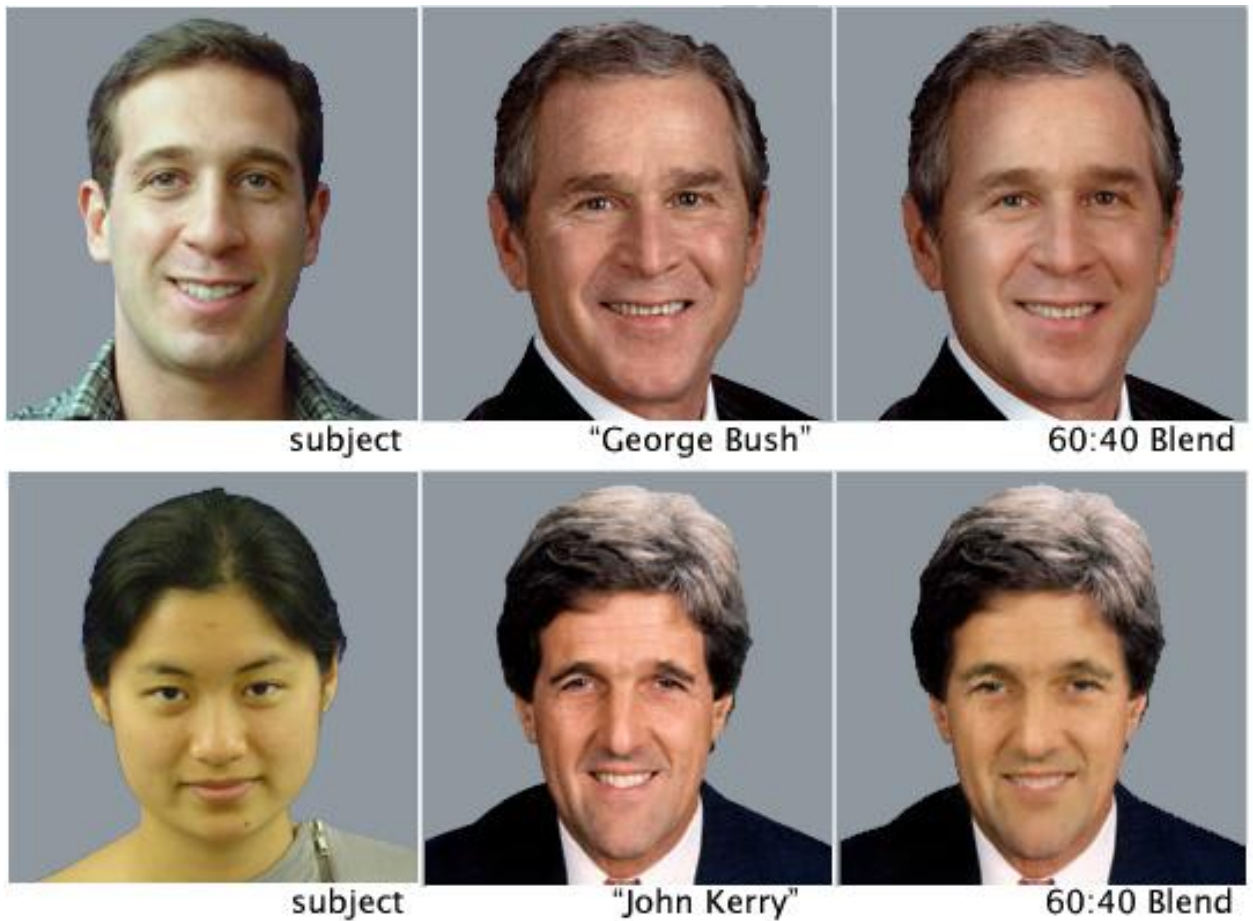
Several lines of research suggest that physical similarity in general, and facial similarity in particular, is a relevant criterion for choosing between candidates. Thus, frequency of exposure to any stimulus -- including faces -- induces a preference for that stimulus over other, less familiar stimuli (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 2009). 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.⁸ Figure 2 shows two of the morphs used in this study.

⁸We 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.

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 2009). Thus, the evidence suggests that non-verbal cues can

influence voting, even in the most visible and contested of political campaigns.⁹

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.

Not only does the experiment provide the most convincing basis for causal inference, experimental studies are also inherently replicable. The same experimental design can be administered independently by researchers in varying locales with different stimulus materials and subject populations. Replication thus provides a measure of the reliability or robustness of experimental findings across time, space, and relatively minor variations in study procedure.

Since the first published reports on the phenomenon of media “priming” -- the tendency of experimental participants to weigh issues they have been exposed to in experimental treatments more heavily in their political attitudes -- the effect has been replicated repeatedly. Priming effects now apply to evaluations of public officials and governmental institutions, to vote choices in a variety of electoral contests, to stereotypes, group identities, and any number of other attitudes. Moreover, the finding has been observed across an impressive array of political and media systems (for a recent review of priming research, see Roskos-Ewoldsen, Roskos-Ewoldsen and Carpentier 2005).

⁹ 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.

The Issue of Generalizability

The problem of limited generalizability, long the bane of experimental design, is manifested at multiple 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 designs in which the intervention is non-obtrusive and the settings more closely reflect ordinary life.⁴

One approach to increasing experimental realism is to rely 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 the treatment 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 home television viewing environment. The fact that the advertising research lab was configured to resemble a typical living or family room setting (complete with reading matter and refreshments) meant that participants did not need to be glued to the television screen. Instead, they could help themselves to cold drinks, browse through newspapers and magazines, or engage in small talk with fellow participants.¹⁰

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

¹⁰ In the early days of the campaign advertising research, the experimental lab included a remote control device placed above the television set. This proved to be excessively realistic as some subjects chose to fast forward the videotape during the ad breaks. The device was removed.

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. Diversity is important not only to enhance generalizability, but also to mount more elaborate tests of mediator or moderator variables. In experiments featuring racial cues, for instance, it is imperative that the study participants include a non-trivial number of minorities. 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.

Since going live in 2001, over 2500 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? First, we gauged the degree of divergence between drop-in participants and typical Internet users. The results suggested that participants in the online experiments reasonably approximated the online user population at least with respect to race/ethnicity, education, and party identification. The clearest evidence of selection bias emerged with age and gender. The mean age of study participants was significantly younger and participants were also more likely to be male. 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 or greater enthusiasm for online games among males.

The second set of comparisons assesses the overlap between our self-selected online samples and all voting-age adults (these comparisons are based on representative samples drawn by Knowledge Networks 2000). Here the evidence points to a persisting digital divide in the sense that major categories of the population remain underrepresented in online studies. In relation to the broader adult population, our experimental participants were significantly younger, more educated, more likely to be white males, and less apt to identify as a Democrat.

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. 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 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.

To date, the Washingtonpost.com – PCL collaborative 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, Li and Chatt 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.¹¹ 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.; Bailey and Rivers 2009).

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. Web experiments eliminate the need for elaborate lab space and resources; all that is needed is a room with a server. These experiments 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

¹¹ The fact that the Polimetrix online samples can be matched according to a set of demographic characteristics does *not* imply that the samples are unbiased. All sampling modes are characterized by different forms of bias and opt-in web panels are no exception. In the US, systematic comparisons of the PMX online samples with RDD (telephone) samples and face-to-face interviews indicate trivial differences between the telephone and online modes, but substantial divergences from the face-to-face mode (see Hill, Lo, Vavreck, and Zaller 2007; Malhotra and Krosnick 2007). In general, online samples appear biased in the direction of politically engaged and attentive voters.

of online research panels has made it possible to administer experiments on broad cross-sections of the American population. All told, these features of web experiments go a long way toward neutralizing the generalizability advantage of surveys.

Although web experiments are clearly a low cost, effective alternative to conventional experiments, they are hardly applicable to all arenas of behavioral research. Most notably, web-based experiments provide no insight into group dynamics or interpersonal influence. Web use is typically a solitary experience and web experiments are thus entirely inappropriate for research that requires placing individuals in some social or group milieu (e.g. studies of opinion leadership or conformity to majority opinion).

A further frontier for web experimentalists will be cross-national research. Today, experimental work in political science is typically reliant on American stimuli and American subjects. The present lack of cross-national variation in the subject pool makes it impossible to contextualize American findings,¹² and also means that the researcher is unable to rule out a family of alternative explanations for any observed treatment effects having to do with subtle interactions between culture and treatment (see Juster et al., 2001). Happily, the rapidity with which public access to the Web has diffused on a global basis now makes it possible to launch online experiments on a cross-national basis. Fully operational online opt-in research panels are already available in many European nations including Belgium, Britain, Denmark, Finland, Germany, the Netherlands, Norway, and Sweden. Efforts to establish and support infrastructure for administering and archiving cross-national laboratory experiments are underway at several universities including the Nuffield Centre for Experimental Social Sciences and the Zurich Program

¹² Indeed, comparativists are fond of pointing out the inherently non-comparative and hence pre-scientific nature of research in American politics.

in the Foundations of Human Behavior.¹³ I suspect that by 2015, it will be possible to deliver online experiments to national samples in most industrialized nations. Of course, given the importance of economic development to web access, cross-national experiments administered online -- at least in the near term -- will be limited to the “most similar systems” design.

In closing, it is clear that information technology has removed the traditional barriers to experimentation in political science, including the need for lab space, convenient access to diverse subject pools, and skepticism over the generalizability of findings. The Web makes it possible to administer realistic experimental designs on a world-wide scale with a relatively modest budget. Given the advantages of online experiments, I expect a bright future for laboratory experiments in political science.

¹³ A useful compilation of online experimental labs is available at <http://psych.hanover.edu/research/exponnet.html>

References

- Ansolabehere, Stephen D. 2006. "The Paradox of Minimal Effects," pp. 29-44 in *Capturing Campaign Effects*, eds. Henry Brady and Richard Johnston. Ann Arbor: University of Michigan Press.
- Ansolabehere, Stephen D. and Shanto Iyengar. 1995. *Going Negative: How Political Ads Shrink and Polarize the Electorate*. New York: Free Press.
- Ansolabehere, Stephen D., Shanto Iyengar and Adam Simon. 1999. "Replicating Experiments Using Aggregate and Survey Data." *American Political Science Review*, 93: 901-10.
- Bailenson, Jeremy, Shanto Iyengar, Nick Yee and Nathan Collins. 2009. "Facial Similarity between Candidates and Voters Causes Influence." *Public Opinion Quarterly* 72: 935-61.
- Bartels, Larry. 1993. "Messages Received: The Political Impact of Media Exposure." *American Political Science Review* 87: 267-85.
- Bennett, W. Lance, and Shanto Iyengar. 2008. "A New Era of "Minimal Effects?" The Changing Foundations of Political Communication." *Journal of Communication* 58: 707-31.
- Bositis, David A, and Douglas Steinel. 1987. "A Synoptic History and Typology of Experimental Research in Political Science." *Political Behavior* 9: 263-84.
- Bradburn, Norman M, Lance J. Rips and Stephen K. Shevell. 1987. "Answering Autobiographical Questions: The Impact of Memory and Inference in Surveys." *Science* 236: 157-61.

- Burnstein, Eugene, Christian Crandall, and Shinobu Kitayama. 1994. "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: 773-789.
- 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.
- Druckman, James N., Donald P. Green, James H. Kuklinski, J., and Arthur Lupia. 2006. „The Growth and Development of Experimental Research in Political Science.” *American Political Science Review* 100: 627-635.
- Finkel, Steven E., and John G. Geer. 1998. "A Spot Check: Casting Doubt on the Demobilizing Effect of Attack Advertising." *American Journal of Political Science* 42: 573-95.
- Freedman, Paul, and Kenneth Goldstein. 1999. "Measuring Media Exposure and the Effects of Negative Campaign Ads." *American Journal of Political Science* 43: 1189-208.
- Gaines, Brian J, and James H. Kuklinski. 2008. "A Case for Including Self-Selection Alongside Randomization in the Assignment of Experimental Treatments." Presented at the Annual Meeting of the Midwestern Political Science Association.
- Gilliam, Franklin Jr., and Shanto Iyengar. 2000. "Prime Suspects: The Influence of Local Television News on the Viewing Public." *American Journal of Political Science* 44: 560-73.
- Gilliam, Franklin Jr., Nicholas A. Valentino, and Matthew Beckman. 2002. "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: 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: 6-23.
- Green, Donald P., and Alan S. Gerber. 2003. “The Under-Provision of Experiments in Political and Social Science.” *Annals of the American Academy of Political and Social Science* 589: 94-112.
- Gunther, Barrie. 1987. *Poor Reception: Misunderstanding and Forgetting Broadcast News*. Hillsdale, NJ: Lawrence Erlbaum.
- Heckman, James J. and Jeffrey P. Smith. 1995. “Assessing the Case for Social Experiments.” *Journal of Economic Perspectives* 9: 85-110.
- Hermann, Charles F., and Margaret G. Hermann. 1967. “An Attempt to Simulate the Outbreak of World War I.” *American Political Science Review* 61: 400-416.
- Hill, Seth, Lo, James, Vavreck, Lynn, and John R. Zaller. 2007. “The Opt-in Internet Panel: Survey Mode, Sampling Methodology and the Implications for Political Research.” Unpublished paper, University of California-Los Angeles.
<http://web.mit.edu/polisci/portl/cces/material/HillLoVavreckZaller2007.pdf>, accessed 3 November, 2008.
- Hovland, Carl I. 1959. “Reconciling Conflicting Results Derived From Experimental and Survey Studies of Attitude Change.” *American Psychologist* 14: 8-17.
- Iyengar, Shanto and Donald R. Kinder. *News That Matters: Television and American Opinion*. Chicago: University of Chicago Press.

- Juster, Thomas F. et al., 2001. *Preparing for an Aging World: The Case for Cross-national Research*. Washington DC: National Academy Press.
- Kahn, Kim F., and Patrick J. Kenney. 1999. "Do Negative Campaigns Mobilize or Suppress Turnout? Clarifying the Relationship between Negativity and Participation." *American Political Science Review* 93: 877-90.
- Kinder, Donald R., and Thomas R. Palfrey. 1993. *Experimental Foundations of Political Science*. Ann Arbor: University of Michigan Press.
- Lau, Richard R., Lee Sigelman, Caroline Heldman, and Paul Babbitt. 1999. "The Effects of Negative Political Advertisements: A Meta-Analytic Assessment." *American Political Science Review* 93: 851-75.
- Malhotra, Neil, and Jon A. Krosnick. 2007. "The Effect of Survey Mode and Sampling on Inferences about Political Attitudes and Behavior: Comparing the 2000 and 2004 ANES to Internet Surveys with Non-probability Samples." *Political Analysis* 15: 286-323
- Mendelberg, Tali. 2001. *The Race Card: Campaign Strategy, Implicit Messages, and the Norm of Equality*. Princeton: Princeton University Press.
- Orne, Martin T. 1962. "On the Social Psychology of the Psychological Experiment: With Particular Reference to Demand Characteristics and Their Implications." *American Psychologist* 17: 776-83.
- Pierce, John C., and Nicholas P. Lovrich. 1982. "Survey Measurement of Political Participation: Selective Effects of Recall in Petition Signing." *Social Science Quarterly*, 63: 164-71.
- Price, Vincent, and John R. Zaller. 1993. "Who Gets the News? Alternative Measures of News Reception and Their Implications for Research." *Public Opinion Quarterly* 57: 133-64.

- Prior, Markus. 2003. "Any Good News in Soft News? The Impact of Soft News Preference on Political Knowledge." *Political Communication* 20: 149-72.
- _____. 2007. *Post-Broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Elections*. New York: Cambridge University Press.
- Riker, William H. 1967. "Bargaining in a Three-Person Game." *American Political Science Review* 61: 642-656.
- Rivers, Douglas R., and Delia Bailey. 2009. "Inferences from Matched-Samples in the 2008 U.S. National Elections." Proceedings of the Survey Research Methods Section of the American Statistical Association, 627-39.
- Rivers, Douglas R. 2006. 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.
- Roskos-Ewoldsen, David, Beverly Roskos-Ewoldsen, and Francesca R. Carpentier. 2005. "Media priming : A Synthesis," in *Media Effects : Advances in Theory and Research*. Ed Jennings Bryant and Dolph Zillmann. Hillsdale, NJ : Lawrence Erlbaum Associates.
- Scioli, Frank. 2009. Personal communication.
- Sears, David O. 1986. "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: 515-30.

- Vavreck, Lynn. 2007. "The Exaggerated Effects of Advertising on Turnout: The Dangers of Self-Reports." *Quarterly Journal of Political Science* 2: 325-43.
- Wattenberg, Martin P., and Craig L. Brians. 1999. "Negative Campaign Advertising: Demobilizer or Mobilizer?" *American Political Science Review* 93: 891-900.
- Zajonc, Robert B. 2001. "Mere Exposure: A Gateway to the Subliminal." *Current Directions in Psychological Science* 10: 224-228.
- Zaller, John R. 1992. *The Nature and Origins of Mass Opinion*. New York: Cambridge University Press.