18

Online Panels and the Future of Political Communication Research

Shanto Iyengar* and Lynn Vavreck

A major objective of political communication research is to assess the effects of different forms of media programming on the voting public. In the 1970s and 1980s, 'media effects' researchers concentrated on television news because that was where most Americans got their news (see Iyengar and Kinder, 1987). As new media and alternative forms of news programming gained audience share, researchers have expanded their focus accordingly.

Despite the profound changes in the media environment and related changes in the content of campaign communications, the fundamental challenges facing media effects researchers remain constant. The most basic is the challenge of obtaining accurate and precise measures of exposure to media messages. Self-reports on media exposure – long the preferred measure – are likely to yield highly exaggerated and hence inaccurate estimates of media effects. Researchers must also deal with the thorny issue of endogeneity (or two-way causation) between exposure to media programming and political predispositions, which makes it difficult to disentangle the causal effects of messages on the opinions and actions of the audience and vice versa.

Problems of imprecise measurement and weak causal inference have proven intractable in observational research such as surveys. In recent years, political communication researchers have increasingly turned to experimental methods that substitute actual exposure to media messages for self-reported exposure, and which neutralize the issue of endogeneity by assigning individuals to particular audiences on a random basis. The combination of manipulational control and random assignment yields unequivocal causal evidence. Can the results of experimental research be generalized to the real world? The standard argument against experimentation is that the findings are of questionable generalizability because of the artificiality of the experimental setting and the unrepresentativeness of the subject pool. We demonstrate, however, that this conventional wisdom no longer applies, at least to experiments administered online. The revolution in information technology enables researchers to overcome the two prime limitations of experimental research. As people become more adept with technology and spend increasing amounts of time with web-based media, experiments administered online become highly realistic for the simple reason that they mirror individuals' actual media experiences. Further, the development of relatively low cost, but probability-based techniques for sampling from nationwide opt-in Internet panels has enhanced both the power and generalizability of web experiments. Thus, technology has not only transformed the way in which candidates and voters communicate with each other, but has also redefined the methodology of political communication research.

THE CHALLENGES FACING OBSERVATIONAL RESEARCH

In March of 2008 – in the midst of one of the most heated presidential primary campaigns in history – Howard Kurtz, the media columnist for *The Washington Post*, asserted that *Saturday Night Live* (SNL) had changed the course of the 2008 presidential election. SNL writers had taken full advantage of the campaign using newsworthy and amusing moments from the candidates' activities (for example, Sarah Palin's appearance on the CBS Evening News) as fodder for their skits and impersonations. Whether it was Amy Poehler spoofing Hillary Clinton or Tina Fey playing Sarah Palin, the show's real-time ratings soared. Moreover, according to Integrated Media Measurement Incorporated (IMMI), twice as many people watched recordings of the campaign sketches – either on a digital video device or on YouTube and other websites – than watched the live broadcast (Welsh, 2008). More interestingly, in the week after the Vice Presidential debate in St Louis, IMMI reported that only half the people who saw an SNL skit on the debate watched debate itself. This disparity in audience size suggests that for many people, the SNL re-enactment was their *only* exposure to the debate.

At the very least, the evidence suggests that there was considerable public exposure to the SNL treatment of the presidential campaign. With so many people watching, Kurtz may have had a point: perhaps the SNL audience's impressions of the candidates were shaped by the programming they encountered on SNL. But can we test Kurtz's assertion more systematically?

To properly isolate the impact of SNL on public opinion, any study design must meet two requirements. First, there must be data on actual exposure to SNL content. Second, the researcher must be in a position to assess the effects of exposure to SNL on candidate preference or vote choice independent of the effects of vote choice on the decision to watch SNL. As described below, relatively few observational studies meet both these requirements.

Measurement

Exposure is a necessary condition for media influence. How can we reliably differentiate people who saw the SNL skits from those who did not? Traditionally, political communication researchers have relied upon survey respondents' self-reports to measure exposure. For instance, respondents might be asked how many times they watched SNL during the past month. Unfortunately, there is considerable evidence that self-reports of media exposure lead to profound biases in our estimates of media effectiveness, exaggerating the impact by as much as 600% (Ansolabehere and Iyengar, 1994; Price and Zaller, 1993; Vavreck, 2007).

To illustrate the problem, we present data from Vavreck's 2002 experimental study of advertising and turnout. Vavreck randomly assigned half of her sample to receive advertising treatments, including a non-partisan get-out-the-vote ad aimed at stimulating turnout in the 2002 midterm elections. The control group saw nothing. Each respondent's registration and turnout record was located after the election from official statewide databases. In Table 18.1, we re-analyze these data to demonstrate the differences in estimates of the treatment effect on actual turnout or self-reported turnout depending on whether we use self-reported or actual exposure to the treatment ad. The true effect of the ad (the effect of actual exposure on actual turnout) is to increase turnout by less than 1 point. But, when we rely on whether people say they recalled seeing an advertisement encouraging them to vote in the election, the effectiveness of the ad on their reported turnout goes up by a factor of 8.

In 'The Exaggerated Effects of Advertising on Turnout: The Dangers of Self-Reports', Vavreck (2007) shows that it is not just forgetfulness among people who were assigned to the ad that makes this difference.[1] In fact, false recall among the people in the control group who could not

Table 18.1 The effects of ads on turnout using self-reports and actual treatment

	Definitely did not see ad/self-reported turnout	*Definitely saw ad/ self-reported turnout*	*Actual control/ validated turnout*	*Actual treatment/ validated turnout*
Did not vote	70.56	63.24	69.80	68.89
Voted[a]	29.44	36.76	30.20	31.11
Total	100 (394)	100 (612)	100 (202)	100 (225)
	Difference = 7.32		Difference = 0.9	

Note: [a]Dependent variable (voted or not) is self-report for columns one and two. The results in columns three and four use validated vote from the files of the Secretary of State.

have seen the ad contributes more to the inflation of the treatment effect in the self-report analysis.

In order to understand how exposure to political communication affects people, we need to know who, in fact, is exposed. Survey researchers have long argued that better questions will reduce the margin of error in self-reports (Althaus and Kim, 2008; Prior, 2009), but as described below, technological advances may make survey questions entirely unnecessary.

In 2006, Jackman, Lewis and Vavreck worked with IMMI, a media measurement company based in San Mateo, California, to track the production of and exposure to campaign advertisements in the New York media market during the 2006 US Senate races in New York and New Jersey.[2] In New Jersey, the campaign was somewhat competitive between Thomas Kean, Jr, and Robert Menendez (the incumbent); in New York the race was entirely one-sided between incumbent Hillary Clinton and challenger John Spencer.[3]

IMMI has leveraged advances in mobile phone technology and Internet-based research panels to track advertising exposure through cellular phones equipped with proprietary software. In exchange for a new mobile phone and highly subsidized phone service, panelists are required to make the new phone their primary mobile device, keep the battery charged and carry the phone with them at all times. Once turned on, phones begin to digitally encode ambient audio on a regular basis (10 seconds of every 30 seconds throughout the entire day). These recordings are reduced to digital 'fingerprints', which are uploaded continuously to the IMMI servers. The phones do not record audio, nor can the digital fingerprint be 'played' at any point in time. The sole purpose of the digital code is to serve as a target that will be matched against a universe of other digital codes that were aired on all media that day in that market. In order to identify the audio fingerprints being sent by the phones, IMMI also tracks all media broadcast in the target market and maintains client-provided content files, such as commercials, promos, movies and songs. By automatically comparing the uploaded audio fingerprints captured by the phones with audio fingerprints stored on the IMMI servers, IMMI identifies all the media any given panelist is exposed to on a daily basis.

This kind of continuous data collection is new to political science and communication studies – the resulting collection is not a static file but instead a data stream that has to be captured and reduced to something useful to researchers. As data from sensing devices such as the IMMI phone become more widely available, we will have to develop new measurement strategies and methods aimed at managing and analyzing these large streams of real-time observations.

The IMMI data, coupled with survey data on any given panelist's background, allow a researcher to investigate who is exposed to advertising, who is affected by exposure, and what determines both of these things. Obviously, continuous tracking of advertising exposure is a major improvement over self-reports that measure exposure over vaguely defined periods of time.

Since the technology measures exposure to all forms of media programming, it is possible to study the effectiveness of political ads or political shows such as SNL, conditional on a whole host of other factors about the media environment, such as the amount of news programming the person was exposed to or the content of the entertainment programming encountered on any particular day. In Figure 18.1, we present data from Jackman et al.'s (2006) analysis of the 411 IMMI panelists who reside in the New York media

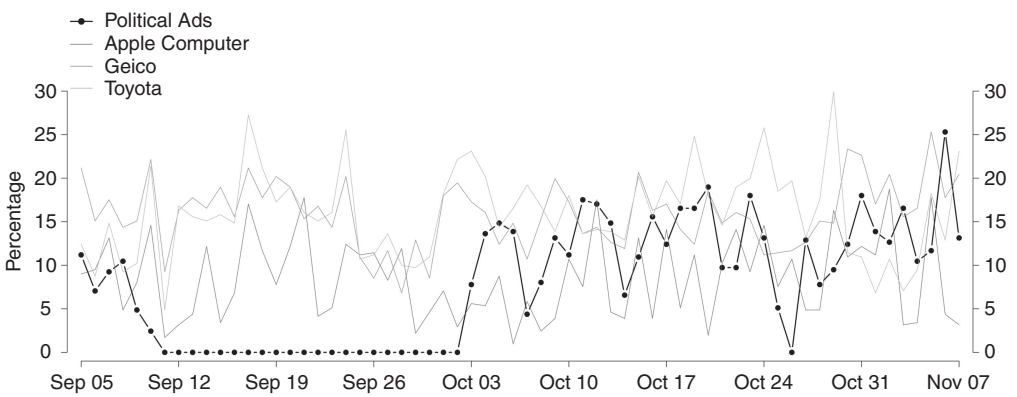

Figure 18.1 Percentage of 411 IMMI panelists viewing at least one ad per day, by selected advertisers, New York DMA

market. The figure demonstrates the relative prominence of political advertising compared to other forms of advertising. Although we tend to think of political advertising as being ubiquitous during campaigns, the data show that in fact, it is not until the last week of the campaign that political ads outnumber ads from Toyota, Geico and Apple – at least in terms of the ads people actually see.

The IMMI data can also be examined by advertiser and day part to show when and where advertisers air their ads. Jackman et al. were thus able to map the different advertising strategies of the candidates over the course of the campaign. Figure 18.2 presents these data for the New Jersey race between Keane and Menendez. Hollow dots represent the airing of an ad, while solid dots indicate that at least one IMMI panelist saw that ad at that time. Clearly, the two campaigns achieved very different levels of exposure to their advertising. Keane had 1614 unique ad buys that were seen 2770 times by 225 unique panelists. Menendez, however, had only 585 buys seen 495 times by 121 people.

What is especially interesting about the data in Figure 18.2 is that volume of advertising is *negatively* correlated with the election outcome. Keane lost the election to Menendez, who ran many fewer ads, exposed only half as many unique people, and whose ads were seen with only one-sixth the frequency of the Keane ads. These data provide a telling illustration of a recurring problem that confronts political campaign research – candidates' choices about when and how much to advertise are endogenous to the outcome we care about, namely, vote choice. In New Jersey, Keane was forced to out-advertise Menendez because he knew he was behind in the polls and had to reach a lot of voters in order to catch up. His advantage in advertising was affected by the state of public opinion rather than the other way around.

Political communication research of the future will increasingly feature technology like that developed by IMMI. As we learn more about developing and managing research panels made up of people who are frequently online or on their mobile phones, we will find ourselves awash in behavioral data about the particular media content to which people are exposed on a given day, which represents quite a dramatic contrast from self-reported media exposure.

Endogeneity and Power

The New Jersey Senate advertising above illustrated one particular version of the problem of endogeneity or reciprocal causation. The problem occurs at two levels. As seen above, candidate behavior is affected by their beliefs about the state of the race. This makes it difficult to tease out the effects of the candidates' advertising on voters' preferences. To make matters worse, the decisions that people make about what particular media programming to watch or listen to are also endogenous to their political preferences. To take an obvious example, Republicans are much more likely than Democrats to watch Fox News. Given SNL's constant lampooning of Sarah Palin, we might expect an equally significant partisan divide in the SNL audience during the 2008 campaign. On both the supply and demand side, therefore, the political communication equation is fraught with endogeneity concerns.

To return to Howard Kurtz's assertion about the transforming role of SNL, evidence from the 2007 wave of a 2007–08 panel study shows that people who report watching political satire shows like SNL are 30% more likely to be interested in politics than those who do not watch these shows – and this is before the campaign heats up (Jackman and Vavreck, 2008). In fact, among people who report exposure to programs like SNL in 2007, more than 66% report being very interested in politics and current affairs – well before these shows began to caricature the candidates. Fans of political satire are also significantly more likely to read political blogs on the Internet and email friends, coworkers and family members about politics. They are more likely to be young, white and male; to identify as Democrats or independents, and to describe themselves as more liberal than conservative (Jackman and Vavreck, 2008). Although this pattern suggests that interest in politics and a preference for Obama motivates attention to SNL political skits, not vice versa, Jackman and Vavreck's data also show that watching SNL in 2007 increased people's chances of voting for Obama in the general election.

These data make it clear why it is so difficult to test Kurtz's assertion with observational data. There is something about a person that makes them more likely to watch shows like SNL, but which also makes her more likely to vote for Obama. Unless we can measure that thing, it will always look like watching the show is driving vote choice or vice versa. The problem is that we rarely know what critical concepts we are not measuring when we write survey questions – and even if we did know, our ability to measure these concepts is crude at best.

So, to know whether Kurtz had it right, we need to know who saw the SNL skits and when they saw them. We need to acquire lots of information about these people so we can account for how their individual characteristics affected whether they viewed the show, and we also need to know their political attitudes and vote intentions – and

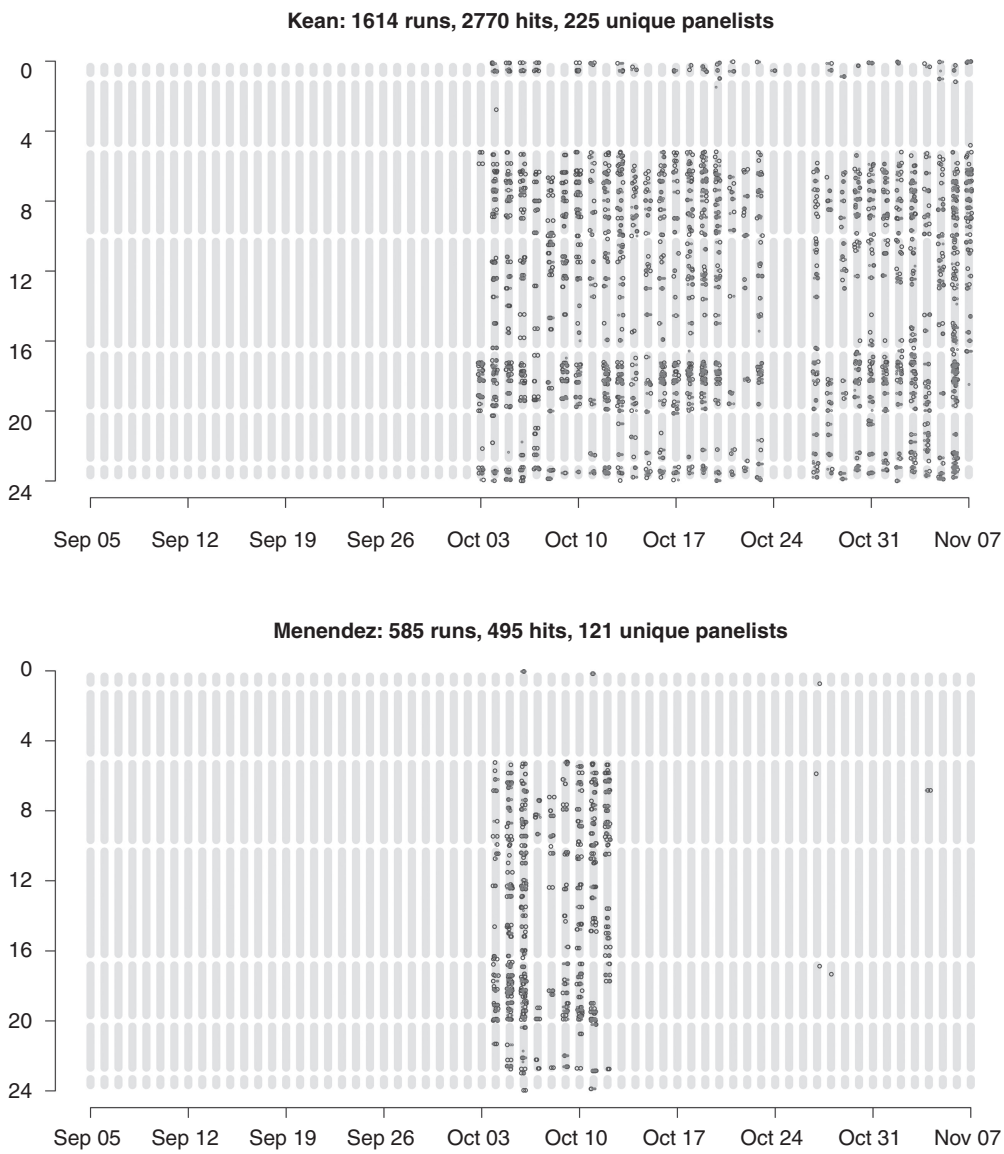

Figure 18.2 New Jersey Senate race advertising profiles, 2006

ultimately whether they voted in the election and for whom. And, we need to know this about tens of thousands of people over several repeated weeks of the campaign so we can eventually uncover what is likely to be a very small effect of SNL on turnout or vote choice. To answer this question we need better measures than we currently have, a research design that gets around the endogeneity issue, and access to larger samples than we typically have to isolate potentially small, but important campaign effects.

On both issues described above – measurement of exposure and endogeneity – observational work is clearly lacking. Survey data are typically based on self-reports and limited to one point in time. Lacking both a precise indicator of just who is in the audience at any given moment and the ability to separate cause from effect, observational methods give us very little leverage over questions of media effectiveness. The more promising solution lies in experimental methodology. However, before moving to experimentation, it is important

to point out that there is one area of survey research in which we have made great strides recently and will continue to make progress in the years to come – and that is in the development of online surveys.

The large-scale diffusion of the Internet has made survey research more affordable – roughly two orders of magnitude less expensive than face-to-face interviewing. For the same amount of money, we could interview 1000 people in person or 86,000 people online.[4] With the significantly lowered per respondent cost, it becomes possible to track opinion over long periods of time for a set of impaneled people, or track people every day for a few weeks at a time. The substantial increase in sample size and the ability to track the same survey respondents over prolonged periods of time help observational researchers gain traction on the endogeneity problems described above.

THE BENEFITS OF EXPERIMENTATION

An alternative approach to the question posed by Howard Kurtz is to administer a randomized experiment in which we expose half the subjects to a few SNL sketches and then determine their voting preferences. 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, that Horton was black? 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 verbally based or in the form of audio-visual cues.

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.

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

Racial Cues in Local News Coverage of Crime

As any regular viewer of television knows, crime is a frequent topic in broadcast news. In response to market pressures, television stations have adopted a formulaic approach to covering crime, an approach designed to maximize audience interest. This 'crime script' suggests that crime is invariably violent and those who perpetrate crime are disproportionately non-white. Because the crime script is encountered so frequently, it has attained the status of common knowledge. Just as we know full well what happens when one walks into a restaurant or airport, we also know – or at least think we know – what happens when a crime occurs (Gilliam and Iyengar, 2000).

In a series of experiments, Gilliam and others documented the effects of both the racial and violence elements of the crime script on audience attitudes (Gilliam and Iyengar, 2000; Gilliam et al., 1996, 2002). For illustrative purposes, we 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, then 'painted' to alter the perpetrator's skin color, and re-inserted into the news report. As shown in Figure 18.3, beginning with two different perpetrators (a white male and a black male), the researchers produced 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, for example, 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 manipulation, these differences in the responses of the subjects exposed to the white or black perpetrators could only be attributed to the perpetrator's race.

Figure 18.3 Race of suspect manipulation

FACIAL SIMILARITY AS A POLITICAL CUE

A consistent finding in the political science literature is that voters support candidates who most resemble themselves on questions of ideology, policy or partisanship. 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 relevant for political choice. In particular, evolutionary psychologists argue that physical similarity is a kinship cue and humans are motivated to treat their kin preferentially (for instance, Burnstein et al., 1994).

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 et al., 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 computer screens on which these candidate questions appeared also 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 the participant.[5] Figure 18.4 shows two of the morphs used in this study.

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 (Bailenson et al., 2009). Thus, the evidence suggests that non-verbal cues can influence voting, even in the most visible and contested of political campaigns.[6]

IMPROVED MEASURES OF AUDIENCE RESPONSE

The ability to launch experiments online has further strengthened the ability of political communication researchers to draw causal inferences

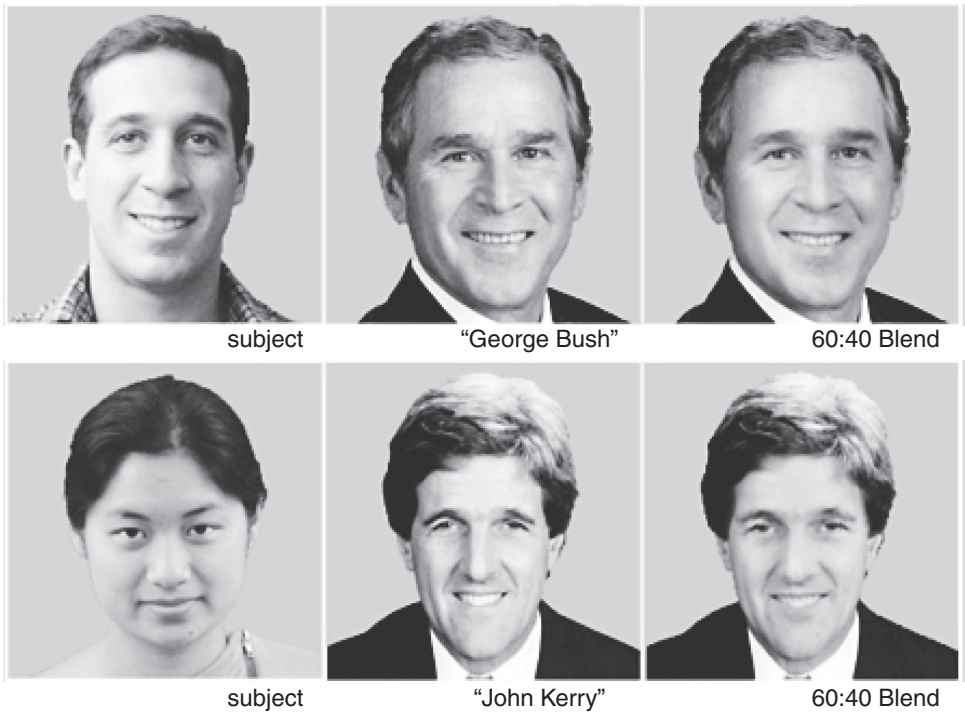

Figure 18.4 Facial similarity manipulation

by providing more precisely calibrated indicators of audience reactions to media messages. Online experiments permit observation of information seeking behavior as well as user reactions to visual, verbal and audio-visual stimuli such as the material encountered in 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 pre-existing beliefs (for example, Festinger, 1957) and instead put themselves in the path of information they expect to agree with. In the words of Lazarsfeld et al:

> 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. (1948: 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. Indeed, Cotton observed that the selective exposure literature had failed to address 'how people actively seek and avoid information on their own' in naturalistic settings (1985: 29). However, 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 1 July and 7 October, 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 as well as the first televised

debate. All told, the information amounted to over 600 pages of text and two hours of multimedia (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 asked 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 (600 people were invited to participate in the study; of these, 226 actually used the CD for a response rate of 38%.) 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 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 News network in 1986, Republicans have enjoyed easy access to television news with a pro-Republican tilt. The tendency to avoid attitude-discrepant information encouraged by Fox News may have promoted similar information-seeking behaviors 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 post hoc summary or 'averaged' assessment of their reaction to the content and imagery of ads.

With the diffusion of online technology, it is possible to monitor viewer response to advertising on a continuous basis, over the entire playing of the ad (Iyengar et al., 2007). Rather than asking for a summary assessment *after* viewers have watched the ad, researchers can use an online 'dial' (or sliding scale) 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.

Ad dial methodology was implemented online in a study of the 2006 US 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 toward 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 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 18.5 displays a screen shot from one of the Tennessee conditions featuring the race between Republican Bob Corker and Democrat Harold Ford Jr, with relatively positive and negative positions of the dial.

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 18.6). This pattern is consistent with prior research showing that exposure to campaign ads strengthens viewers' partisan predispositions (Ansolabehere and Iyengar, 1996). While partisans responded rapidly to the content of advertising, Independents were typically unmoved, remaining lukewarm over the entire playing of the ad.

A further finding from this study was that the rate of polarization proved variable across

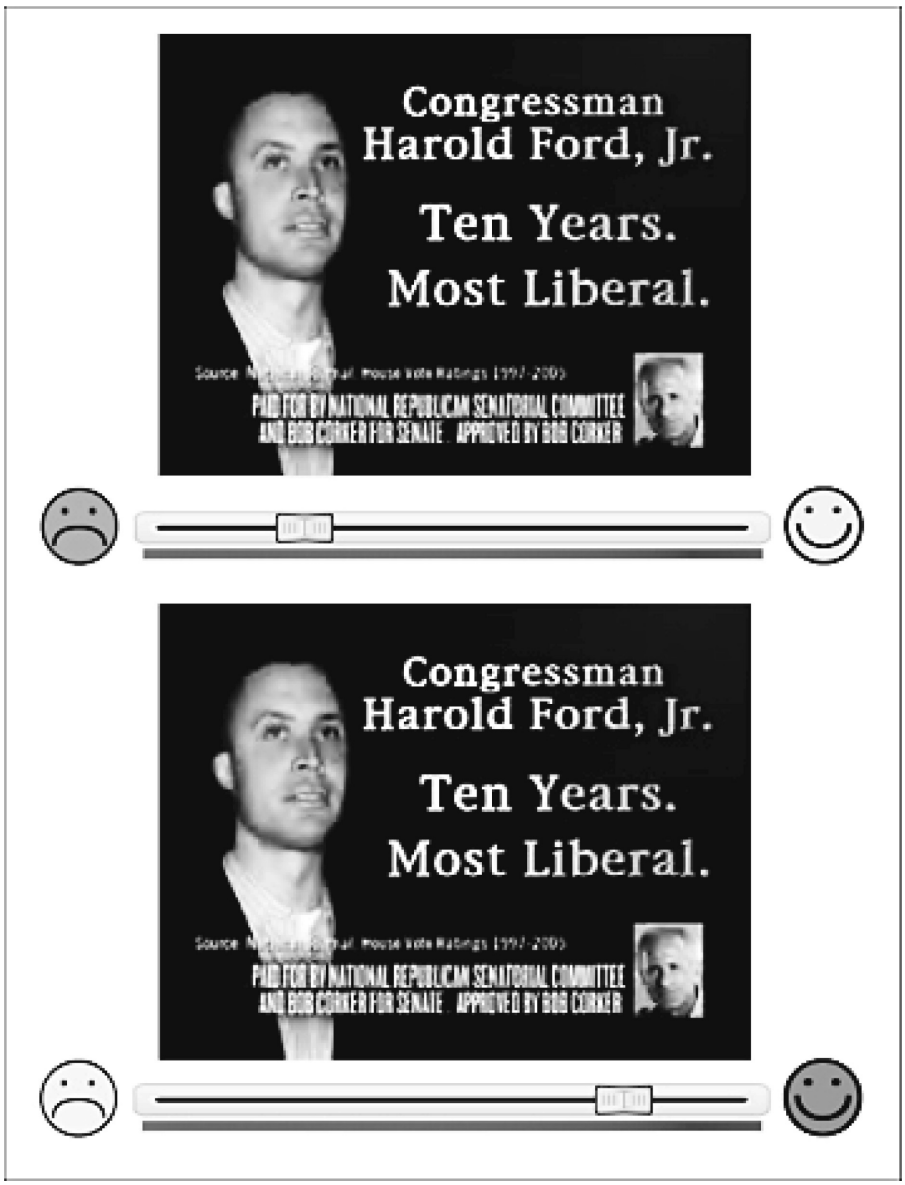

Figure 18.5 Screen shots from online dials
***Note*: As the ad played, participants could move the slider to indicate their feelings about the content of the ad, with the position of the dial recorded once a second.**

the 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 and the fallout from the Abramoff ethics scandal linking the Republican candidate with President Bush, mobilized the Democratic base more effectively than generic Republican appeals on national security, immigration and taxes.

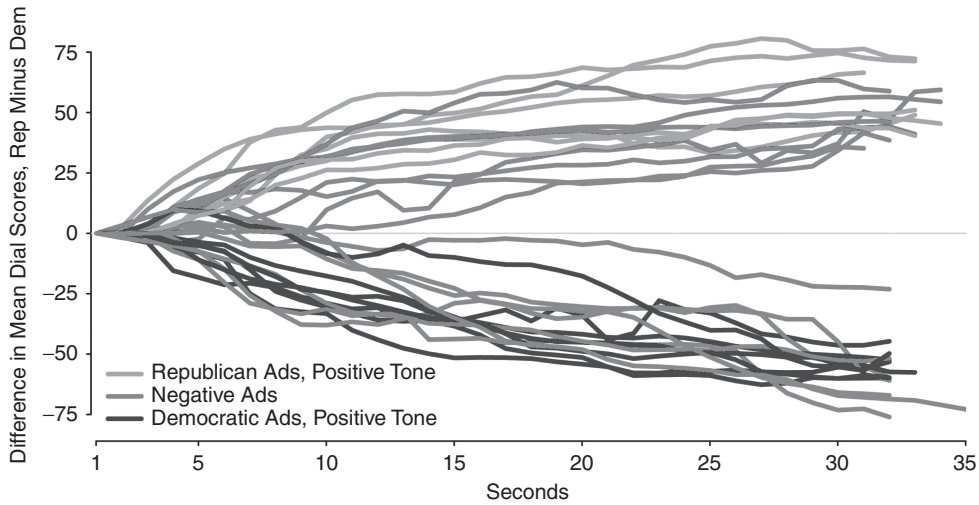

Figure 18.6 Partisan polarization in dial score

THE CHALLENGE FACING EXPERIMENTAL RESEARCH

The problem of limited generalizability, long the Achilles Heel of experimental design, occurs at three different 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 tight control over exposure to the stimulus, the laboratory setting in which the experiment occurs is often quite dissimilar from the setting in which subjects ordinarily experience the 'target' phenomenon. The inherently artificial properties of lab experiments have led researchers to turn to designs in which the procedures and settings more closely reflect ordinary life.

One approach to increased realism is the use of interventions with which subjects are familiar. The Ansolabehere/Iyengar campaign advertising experiments, conducted in the Los Angeles area in the early 1990s, were realistic in the sense that they occurred during ongoing campaigns characterized by heavy levels of televised advertising (Ansolabehere and Iyengar, 1996). The presence of political advertisements in a local newscast (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, editing the treatment ads to be representative of the genre. (The need for experimental 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 for the participant equivalent to 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 material (newspapers and magazines) during the viewing sessions. In many 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 target 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, the researchers resolved the experimental realism versus mundane realism tradeoff by embedding the manipulation in a commercial break of a local newscast. For each condition, the treatment 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 modest, amounting to 30 seconds of a 15-minute recording.

In general, there is a significant tradeoff between experimental realism and manipulational 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 (for example, by including a soundtrack and featuring the sponsoring candidate) would have necessarily introduced a series of confounding 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 they are hardly comparable with '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 the crime news experiments conducted in Los Angeles in the summer and fall of 1999, the total cost per subject amounted to approximately US$45. Fortunately, as we have already noted in the case of survey research, 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. The technology is sufficiently user friendly that most web users can now 'self-administer' experimental manipulations. Compared with conventional shopping mall studies, therefore, the costs of online experiments are minimal. Moreover, with the development of Internet research panels, it is now possible to use more rigorous sampling methods to select participants for experimental research.[7] In this sense, online experiments now emulate surveys in the sense of drawing representative samples from some national population.

Sampling from Online Research Panels

Even though online experiments administered on purely opt-in samples may provide more generalizability than the typical 'college sophomore' sample, the self-selected online participant pool is significantly biased in several important respects. The 'digital divide' is still large enough to create significant differences in socioeconomic status between participants and non-participants in online research. Online experiments that attract participants from particular media websites (such as Washingtonpost.com) are also biased in important respects (such as political ideology). Fortunately, it is now possible to overcome issues of sampling bias (assuming the researcher has access to some minimal level of funding) by administering online experiments to representative samples. In this sense, the lack of generalizability associated with experimental designs can be 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 (KN) based in Menlo Park, and the second is Polimetrix (recently purchased

by the UK polling company of YouGov) based in Palo Alto (PMX).

KN 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 probability sampling methods. At its inception, KN invited over 100,000 people living in randomly selected households to join their panel. The final panel included over 150,000 Americans between the ages of 16 and 85 to whom KN provided free access to the Internet via WebTV. In exchange, panel members agreed to participate (on a regular basis) in research studies being conducted by the firm. Today, the panel is smaller and the recruitment methods have changed, but panelists still take surveys online, either on a KN installed WebTV or on their own computers. Thus, in theory, KN delivers samples that meet the highest standards of probabilistic sampling. In practice, because the opt-in rate at initial invitation is quite low, and because those who do opt-in have an obligation to participate, KN has artificially high with-in panel response rates (Dennis et al., 2004). Still, at the heart of the panel-building process is a probabilistic methodology. KN typically delivers data from this panel that come with post-stratification weights to correct for biases in non-response just as a researcher would do who engaged in a traditional probabilistic sampling routine.

The KN method, because it starts with the probabilistic base, is expensive. To avoid the costs of paying for people's Internet usage, the PMX panel (called *PollingPoint*) is made up entirely of people who already have access to the Internet. Further, PMX does not invite panelists to join *PollingPoint* unsolicited, but instead recruits people into the panel after they have completed one PMX online survey by choice.[8] Because this method of recruitment is low-cost, PMX is able to be constantly recruiting panelists, which means that the *PollingPoint* panel is quite large – well over 1 million people. The size of this pool allows researchers at PMX to use a novel 'matching' approach to turn the purely opt-in pool into representative samples.

To extract a representative sample from this pool of self-selected panelists, PMX engages in a two-step process called sample matching. First, PMX constructs a sampling frame from the American Community Study (ACS) with additional data from the Current Population Survey voter supplement and the Pew Religious Life study. From this frame, PMX draws a stratified random sample of people similar in size to the sample they want to produce from their opt-in panel. This is called the target sample.

The next step is to search within the *PollingPoint* panel for respondents who most closely match the individuals in the target sample. On average, two–three matches are drawn for every person in the target sample and all of these people are invited to complete a client's survey. From this set of completed interviews, the final matched-sample is drawn taking the panelists who most closely match the target individuals. This method literally substitutes members of the PMX panel for individuals in a randomly drawn target sample from a corresponding frame.[9] Rivers (2006) describes the conditions under which the matched sample approximates a true random sample.

Although this sampling method is new to communication and political science, non-random sampling of various sorts has been around for decades. The PMX matched-samples have achieved impressive rates of predictive validity, thus bolstering the claims that matched samples emulate random samples on the criterion of representativeness. In the 2005 California special election, PMX 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, 2009). In 2006 and 2008, PMX correctly predicted senate and gubernatorial elections with minimal bias (Vavreck and Rivers, 2008).

Whether one is more comfortable with the KN approach to sampling that has its roots in a probabilistic method, or the PMX approach, which uses matching and weighting to remove selection biases, the truth of the matter is that the Internet is here to stay. No matter how hard researchers hold on to random sampling methods like RDD or area probability sampling, the future of communication research is online. The fact that many people now get their news, communicate with friends and watch entertainment on their computers only underscores the need for us to make the transition to online research methodologies.

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 we have suggested, however, traditional experimental methods can now be effectively and rigorously replicated using online strategies which have the advantage of reaching a participant pool that is more far-flung and diverse than the pool relied on by conventional experimentalists. The development of online research panels

makes 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 information 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 of available information, people uninterested in politics can avoid news programming altogether while political junkies can exercise discretionary or selective exposure to political information. Thus, random assignment is no longer an entirely appropriate strategy for assigning individuals to audiences (Gaines and Kuklinski, 2008). In other words, as we force subjects in experiments to comply with our treatments, we are assuming that the decision about whether to experience the treatment in the real world is unrelated to the size of the treatment's effect. If the decision about whether to watch an advertisement or read a news story conditions the effect of the treatment, not even new technology can save us from making mistaken inferences about effects.[10]

This increasing degree of endogeneity between the composition of particular media audiences and their political predispositions (and the potentially conditioning effect) has important consequences for the design of experimental research. Two recent projects highlight the importance of clever designs and the use of new technology to explain the role of political communication. Vavreck and Green (2009) take the endogeneity seriously by administering a large-scale field experiment testing the effects of non-partisan get-out-the-vote (GOTV) advertisements. They made and ran ads on popular networks in randomly assigned local cable systems during a real election. By using publicly available voter registration files that link voters to zip codes and thus cable systems, Green and Vavreck (2008) are able to demonstrate that running GOTV ads one week before a presidential election in primetime increases turnout on the target group by roughly 3 points. They are careful to call this effect the 'intent to treat' (ITT) effect, since they do not know exactly who in their treatment zip codes actually saw the ads. The ITT effect may be different from the 'treatment on the treated' (TOT) effect if not everyone who lives in the treated zip codes saw the ads – and surely this is the case. By moving the advertising experiment out of the lab and even off the Internet, Vavreck and Green are able to allow people in the treatment zip codes to 'select' in to being treated by whatever mechanisms they actually use in a real campaign when real campaign ads come on the television.

Similarly, Iyengar and Jackman (2003) randomly assigned young voters in California to receive a CD featuring the candidates contesting the 2002 gubernatorial election. Iyengar and Jackman found that actual (validated) turnout among the voters in the treatment group who used the CD was 10 percentage points higher than among young voters in the control group (who did not receive the CD). Like Vavreck and Green (2009), Iyengar and Jackman allowed people in the treatment condition to decide, on their own terms, whether they would engage with the treatment. Iyengar and Jackman speculated, however, that the large 10-point effect – the treatment on the treated effect – 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 and the determinants of the choice also condition the effect of the treatment, it becomes necessary to estimate the average treatment effect after adjusting for self-selection.

Put more clearly, comparing treatment respondents who used the CD to the entire control group was not the right comparison – even though subjects were assigned randomly. To demonstrate this, the researchers identified a comparison set of 'interested' subjects in the control group who are the closest matches to those interested people in the treatment group who actually used the CD. In the CD experiment, 78% 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 the politically engaged. Thus, actual engagement with the treatment was in fact 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 (Heckman et al., 1998; Imbens, 2004). The general idea is straightforward: although respondents have self-selected into a particular treatment or experimental condition, 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). When the

authors of the CD study used propensity score matching to generate a matched control group for their interested treatment group subjects, their estimate of the effects of the CD on turnout was reduced by more than one-half.

In summary, the use of digital technology in experimental research represents a double-edged sword. Although researchers are in a position to administer precisely controlled manipulations to an increasingly large and heterogeneous subject pool, thus increasing both statistical power and generalizability, they face a radically altered media environment in which exposure to political content is increasingly driven by choice (see Bennett and Iyengar, 2008). As a result, assignment to media treatments in the real world will inevitably depend on the participant's political preferences and estimating the effects of these treatments will require the use of more powerful statistical tools – or different experimental designs.

NOTES

* Iyengar's contribution to this chapter was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2010-330-B00028).

1 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 30 minutes later*, to recall having seen the advertisement (Ansolabehere and Iyengar, 1994).

2 Candidates running for office in New Jersey are forced to advertise on New York City television stations because the broadcasts reach significant areas of New Jersey.

3 The biggest spender in the New York media market was Clinton, whose overall spending was reported at US$35,364,218.00. Spending for Tom Kean, Jr, was US$18,463,780.00; and, Menendez spending was more modest at US$14,364,495.00. John Spencer did not advertise in this DMA at all, but reports overall campaign spending of US$5,660,688.00.

4 This calculation is based on face-to-face interview costs of US$1300 for two waves totaling 70 minutes and Internet interview costs of US$15.00 for two waves totaling 20 minutes. This works out to US$18.75 per minute in person and US$0.75 per minute online. These are roughly the respective costs of the 2008 American National Election Study and the 2008 Cooperative Campaign Analysis Project.

5 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.

6 Facial similarity is necessarily confounded with familiarity – people are familiar with their own faces. There is considerable evidence (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.

7 A challenge to online research is the ability to contact people via the Internet. Unlike phone numbers, there is no collection of email addresses from which researchers can randomly draw. Moreover, anti-spam laws make it difficult for researchers to contact people via email without their express permission to do so in advance of the contact. As a consequence, Internet research companies have to build and maintain panels of people who agree to take surveys online for them. Once people agree to receive emails from an online research company, they become members of that company's panel. Most companies maintain very large panels and reward the panelists for completing surveys. Panel management is an important job at any online research firm, and requires creativity and diligence. The goal is to make the survey experience enjoyable for people, not tedious or bothersome. This requires limiting the length of surveys as well as managing the content to make sure questions are not too long or repetitive and the content is not offensive in some way.

8 In order to do this, PMX writes short, fun surveys about a variety of topics (entertainment, sports, pop-culture and politics, among others) and buys online advertising on related web sites that link to the short survey. Respondents share their opinions about everything from gay marriage to favorite recipes and then they are asked to join the *PollingPoint* Panel.

9 Because no match is perfect, the samples may still miss on some combinations of characteristics, thus PMX provides post-stratification weights. The matched cases are weighted to the sampling frame using propensity scores.

10 This, however, is one of the advantages of field experiments, particularly on advertising, since compliance with the treatment is not forced and researchers can calculate the intent-to-treat effects separately from the treatment-on-the-treated effects.

REFERENCES

Althaus, S. L. and Kim, Y. (2008) 'Priming Effects in Complex Information Environments: Reassessing the Impact of News Discourse on Presidential Approval', *Journal of Politics,* 68(4): 960–76.

Ansolabehere, S. and Iyengar, S. (1994) 'Messages Forgotten: Mis-reporting in Surveys and the Bias Toward Minimal Effects', unpublished paper. Department of Political Science, UCLA.

Ansolabehere, S. and Iyengar, S. (1996) *Going Negative*. New York: Free Press.

Bailenson, J., Iyengar, S., Yee, N. and Collins, N. (2009) 'Facial Similarity between Candidates and Voters Causes Influence', *Public Opinion Quarterly*, 72(5): 935–61.

Burnstein, E., Crandall, C. and Kitayama, S. (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(5): 773–89.

Bennett, W. L. and Iyengar, S. (2008) 'A New Era of Minimal Effects? The Changing Foundations of Political Communication', *Journal of Communication*, 58(4): 707–31.

Cotton, J. (1985) 'Cognitive Dissonance in Selective Exposure', in D. Zillman and J. Bryant (eds.), *Selective Exposure to Communication.* Hillsdale, NJ: Lawrence Erlbaum Associates. pp. 11–34.

Dennis, J., Li, R. and Chatt, C. (2004) 'Benchmarking Knowledge Networks' Web-Enabled Panel Survey of Selected GSS Questions Against GSS In-Person Interviews', Knowledge Networks Report.

Festinger, L. (1957) *A Theory of Cognitive Dissonance*. Evanston, IL: Row, Peterson.

Gaines, B. and Kuklinski, J. (2008) 'A Case for Including Self-Selection alongside Randomization in the Assignment of Experimental Treatments', paper presented at 2008 Midwest Political Science Association Meeting. Chicago, IL.

Gilliam, F. Jr and Iyengar, S. (2000) 'Prime Suspects: The Influence of Local Television News on the Viewing Public', *American Journal of Political Science*, 44(3): 560–73.

Gilliam, F. Jr, Valentino, N. and Beckman, M. (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(4): 755–80.

Gilliam, F. Jr, Iyengar, S., Simon, A. and Wright, O. (1996) 'Crime in Black and White: The Violent, Scary World of Local News', *Harvard International Journal of Press/Politics*, 1(3): 6–23.

Green, D. and Vavreck, L. (2008) 'Analysis of Cluster-randomized Experiments: A Comparison of Alternative Estimation Strategies', *Political Analysis,* 16(2): 138–52.

Heckman, J., Ichimura, H. and Todd, P. (1998) 'Matching as an Econometric Evaluation Estimator', *Review of Economic Studies,* 65(2): 261–94.

Imbens, G. (2004) 'Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review', *Review of Economics and Statistics,* 86(1): 4–29.

Iyengar S. and Kinder, D. (1987) *News That Matters: Television and American Opinion*. Chicago, IL: University of Chicago Press.

Iyengar, S., Hahn, K., Krosnick, J. and Walker, J. (2008) 'Selective Exposure to Campaign Communication: The Role of Anticipated Agreement and Issue Public Membership', *Journal of Politics*, 70(1): 186–200.

Iyengar, S. and Jackman, S. (2003) 'Technology and Politics: Incentives for Youth Participation', presented at the International Conference on Civic Education Research.

Iyengar, S., Jackman, S. and Hahn, K. (2007) 'Polarization in Less than 30 Seconds: Continuous Monitoring of Voter Response to Campaign Advertising', presented at the Annual Meeting of the American Political Science Association.

Jackman, S. and Vavreck, L. (2008) *The Cooperative Campaign Analysis Project.* Palo Alto and Los Angeles, CA.

Jackman, S., Lewis, J. and Vavreck, L. (2006) 'Digital Fingerprints: Tracking Ad Exposure in Senate Elections', working paper. UCLA.

Lazarsfeld, P., Berelson, B. and Gaudet, H. (1948) *The People's Choice: How the Voter Makes up his Mind in a Presidential Campaign.* 2nd edn. New York: Columbia University Press.

Price, V. and Zaller, J. (1993) 'Who Gets the News? Alternative Measures of News Reception and Their Implications for Research', *Public Opinion Quarterly,* 57(2): 133–64.

Prior, M. (2007). *Post-Broadcast Democracy: How Media Choice Increases Inequality in Political Involvement and Polarizes Elections.* New York: Cambridge University Press.

Prior, M. (2009). 'Improving media effects research through better measures of news exposure', *Journal of Politics,* 73(1): 893–908.

Rivers, D. (2006) 'Sample Matching: Representative Sampling from Internet Panels', http://www.polimetrix.com/documents/Polimetrix_Whitepaper_Sample Matching.pdf

Rivers, D. (2009) 'Scientific Sampling for Online Research', http://www.polimetrix.com/documents/Polimetrix_Sampling.pdf

Rosenbaum, P. and Rubin, D. (1983) 'The Central Role of the Propensity Score in Observational Studies for Causal Effects', *Biometrika*, 70(1): 41–55.

Sears, D. (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(3): 515–30.

Vavreck, L. (2007) 'The Exaggerated Effects of Advertising on Turnout: The Dangers of Self-Reports', *Quarterly Journal of Political Science*, 2(4): 325–43.

Vavreck, L. and Green, D. P. (2009) 'Rocking the Vote: A Randomized Field Test of Advertising Effectiveness', working paper. UCLA.

Vavreck, L. and Rivers, D. (2008) 'The 2006 Cooperative Congressional Election Study', *Journal of Elections, Public Opinion, and Parties,* 18(4): 355–66.

Welsh, A. (2008) 'SNL's Palin Sketches Turn TV Upside Down: Draw Huge After-Air Audience', integrated Media Measurement Incorporated (IMMI) working paper.

Zajonc, R. (2001) 'Mere Exposure: A Gateway to the Subliminal', *Current Directions in Psychological Science,* 10(6): 224–28.