

CHAPTER 5

CONTENTS

■ What Is an Experiment?	118
■ Why Control Means Stronger Inferences	119
■ Types of Experiments	122
■ Designing the Experiment	136
■ Analyzing and Presenting Results from an Experiment	138
■ Avoiding Mistakes in Your Experiment	141
■ Natural Experiments	143
■ Conclusion	147
■ Key Terms	148

Experiments

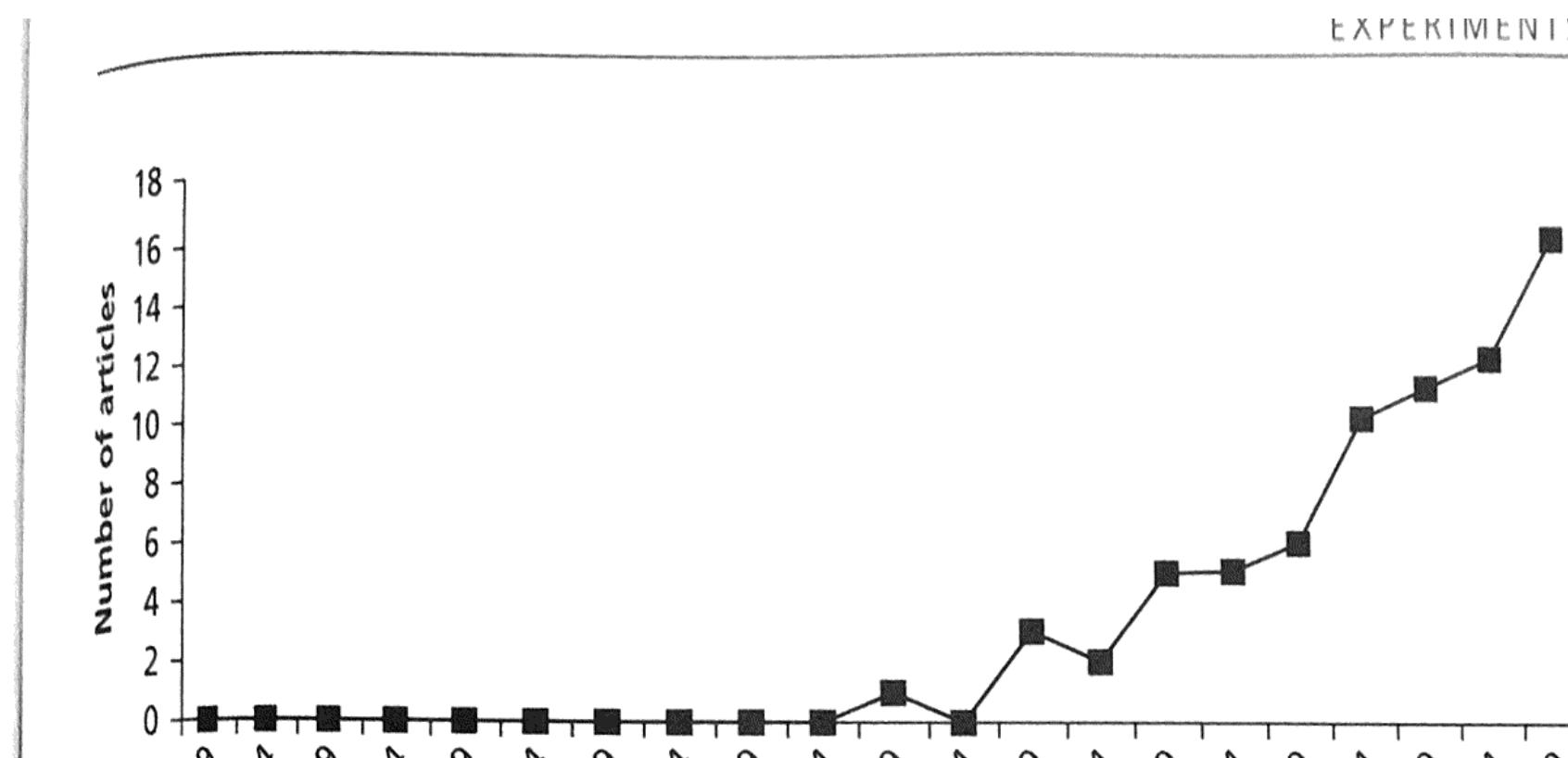
Chances are, an experiment was one of the first research designs you ever executed. Most primary school students in America are asked by their science teachers to conduct an experiment of some kind, with the results presented at science fairs across the country. These experiments may have involved exposing the same types of plants or seeds to different types of light, different temperatures, or different types of soil and then observing whether they grow at different rates. The point of the exercise was not to generate some new scientific finding, but rather to teach you how to use the scientific method to better understand how the world works.

Ironically, despite the fact that they were introduced to the method early in their lives, most college students do not even consider conducting an experi-

ment to learn something about politics. In fact, over the years, few political scientists have considered this method either. Figure 5.1 tracks the percentage of articles published in the *American Political Science Review* (APSR, the discipline's top journal) that used an experiment of some kind. Note that prior to the 1990s, experimental work was quite rare. Prior to 1975, only five articles that appeared in the journal made use of an experiment. Between 1975 and 1990, another sixteen articles in the APSR produced findings from an experiment.

However, political scientists have increasingly begun to recognize the possibilities that experiments hold for answering political questions.¹ During the

1 An alternative interpretation of these figures would be that the top journals have simply become more willing to publish experimental work than they had been previously. But a comparison of papers presented at political science conferences suggests that this is not the case. More political scientists are doing experiments now than ever before.



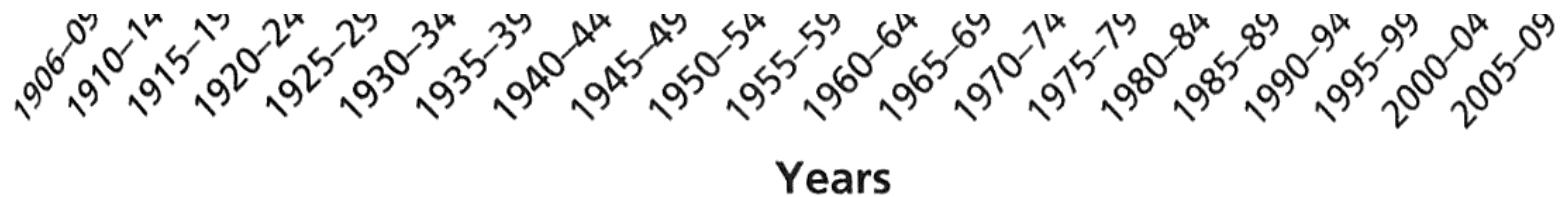


FIGURE 5.1 Articles in the *American Political Science Review* Using Experimental Methods

Source: Druckman, Green, Kuklinski, et al. (2011).²

1990s, twenty-one articles published in the APSR utilized experiments, about twice as many as during the 1980s. And during the first decade of this century, experimental research was used in twenty-eight articles published in the top journal. To put it simply, the experimental method has graduated from obscurity in the discipline and is now extolled as the “gold standard” of research methods by many methodologists.³ As political scientist Rose McDermott notes, “No other methodology can offer the strong support for the causal inferences that experiments allow.”⁴

Why has experimental research made this transition so quickly and emphatically? One reason is that experiments are now easier to conduct than they have been in the past. As we discuss below, the ability to carry out experiments on computers and over the internet has greatly increased access to the experimental method from a wide array of researchers, including student researchers. Another reason for the increasing use of experiments is that political scientists are increasingly collaborating with scholars in other disciplines, such as psychology and economics, which have used these techniques for years. Finally, after decades of advancements in statistical methods designed to draw stronger inferences from observational data, several political scientists have actively promoted the expansion of experimental methods in the discipline. This movement largely focuses on the notion that experiments (and natural experiments) hold enormous potential for allowing political scientists to draw stronger inferences about causal relationships in the political world.

- 2 James N. Druckman, Donald P. Green, James H. Kuklinski, *et al.*, eds, *Cambridge handbook of experimental political science* (Cambridge, UK: Cambridge University Press, 2011).
- 3 Donald P. Green and Alan S. Gerber, “The underprovision of experiments in political science,” *Annals of the American Academy of Political and Social Science* 589 (2003): 94–112.
- 4 Rose McDermott, “Experimental methodology in political science,” *Political Analysis* 10 (2002): 325–342, p. 38.

EXPERIMENTS

WHAT IS AN EXPERIMENT?

An experiment is generally defined by the researcher’s control over what is called the **data-generating process**.⁵ In observational research, as will be discussed in the next two chapters, the researcher does not affect the generation of the data; she simply observes whatever data is produced, however it is produced. In an experiment, the researcher manipulates some variable herself, thereby directly intervening in the production of the relevant data. To understand the distinction, let’s take the example of research on the effects of campaign advertisements that appear on television. This topic is one that has generated significant debate both in political science and in other disciplines like communication. Some scholars argue that television advertisements have “minimal effects” on persuading the public, while others suggest that advertisements can be more influential. We will set aside, for now, the question of which side is right and instead discuss two different ways that one might study this question.

If a political scientist was going to take an observational approach to the study of advertising effects, such a study might look something like the following. First, the researcher could collect data on which candidates ran which advertisements in which television markets.⁶ Then, the researcher might go about collecting election results for each of those television markets. One

about collecting election results for each of those television markets. Once the researcher had collected both sources of data, he could examine whether, for example, the Democratic candidate received a greater share of the vote in markets where she ran more advertisements than the Republican. By extension, the researcher would also look to see whether the Republican won more votes in areas where he ran more advertisements than the Democrat. Such a design would be deemed observational because the researcher had no hand in the process that generated the relevant data that is being analyzed. Specifically, candidates, parties, and interest groups made decisions about which advertisements would be aired and where those advertisements would be aired, not the researcher.

An experimental study of advertising effects would look quite different. Specifically, the researcher might take a variety of different approaches in order to exercise some control over how the data are generated. For example, a researcher might work with one or more of the candidates to determine which television markets to show advertisements in. This way, the researcher is helping to determine how the data are generated. Alternatively, the researcher may set up a living room with a television and invite people to watch different

- 5 Rebecca B. Morton and Kenneth C. Williams, *Experimental political science and the study of causality: From nature to the lab* (Cambridge, UK: Cambridge University Press, 2010).
- 6 Indeed, such a dataset exists for elections held in 1996, 2000, 2002, 2004, and 2008 through the Wisconsin Advertising Project (<http://wiscadproject.wisc.edu/>).

L A T E R V I E W S

advertisements. Some participants might see only the Democratic candidate's advertising while others might only see the Republican's ads. The researcher could then ask participants which candidate they would be more likely to

support. In either case, the researcher now has a hand in the process that is producing the data. She is determining which people will see which ads and then looking for the effects of the different conditions after the fact.

It is important to note that the key here is that the researcher has control over some important variable being studied, not just that the researcher is involved in some way. For example, it would not be sufficient for the researcher to help a candidate place orders for television advertisements. Rather, the key distinction would be whether the researcher was able to determine, to some extent, which advertisements would be aired and where they would be aired. This control is crucial for drawing strong inferences, as we shall see.

WHY CONTROL MEANS STRONGER INFERENCES

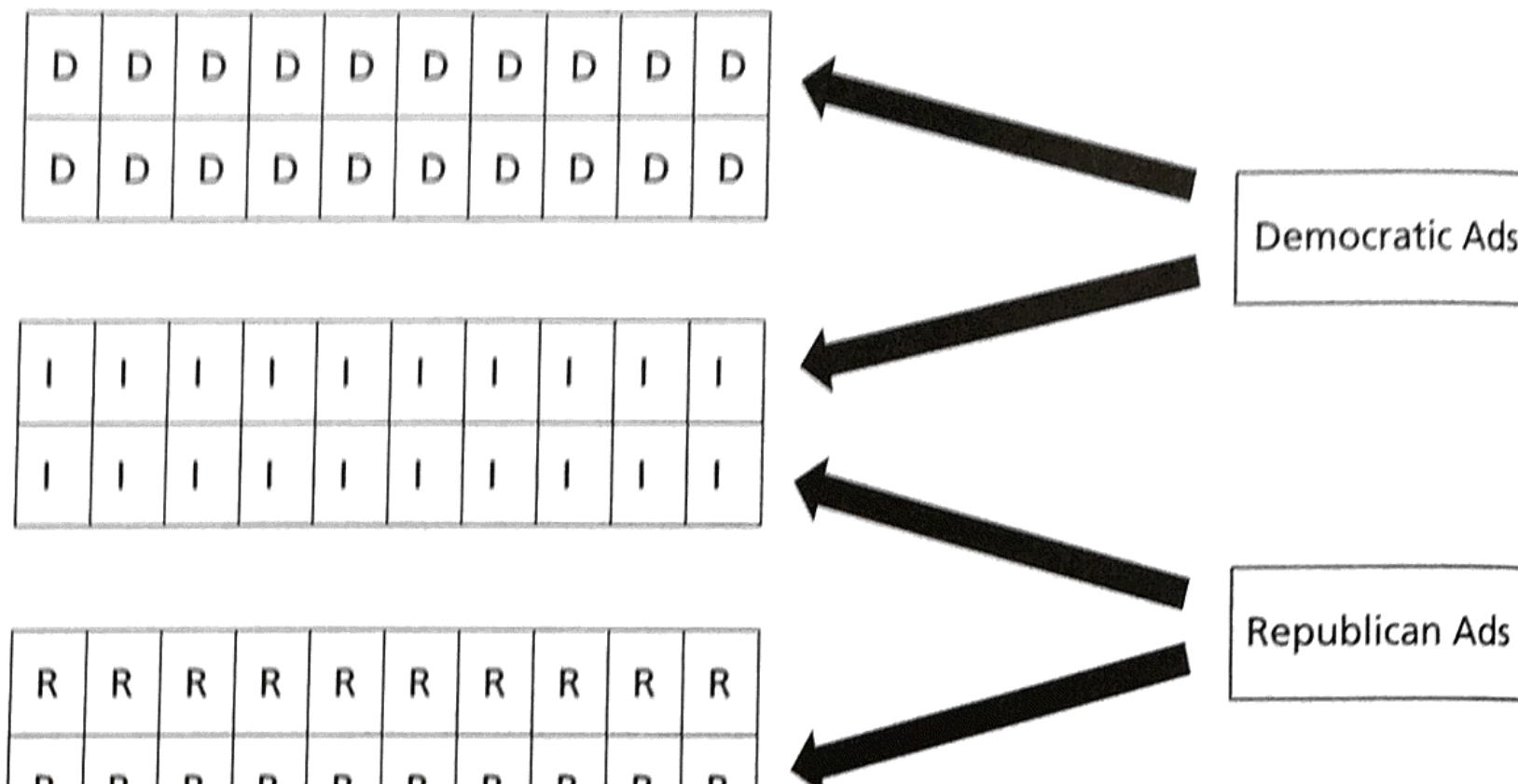
The political world (alas, the world in general) is a complicated place. There are innumerable variables that affect every outcome we observe, whether political or otherwise. For example, an individual's vote may have been influenced by seeing television advertisements, but it will also have been affected by his sense of loyalty to one party or the other, how he has fared economically in recent years, how his friends and relatives are voting, and countless other factors. This matters because to determine whether advertisements influenced his vote, we must isolate this factor and understand its effect separately from the other factors.

To explain this point more clearly, let's imagine a simpler world where the only two factors that affect vote choice are party identification and the advertisements that an individual sees. In such a world, we might be tempted to simply compare whether individuals exposed to more of a particular candidate's advertisements were more likely to vote for that candidate. Whatever difference we found between voters who saw more of the candidate's advertisements compared to those who saw less could then be attributed to the effect of

ments compared to those who saw less could then be attributed to the effect of the advertising. Or could it? The problem, even in this simple world, is that an individual's party identification is likely to affect how much advertising they see. We know, for example, that candidates attempt to show more advertisements to individuals who do not affiliate with either party in order to win over these "swing voters." It is also the case that Democratic candidates tend to show more advertisements to Democratic voters (as a way of encouraging them to turn out to vote), while largely ignoring Republicans. The opposite is true for Republican candidates.

Figure 5.2 presents a representation of how this could interfere significantly with the inferences we would draw in an observational study. This figure shows a hypothetical group of sixty voters—twenty Democrats, twenty

EXPERIMENTS



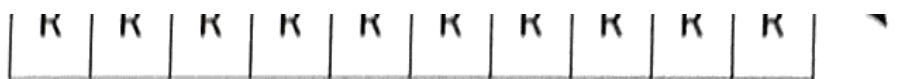


FIGURE 5.2 How Advertising Might Be Distributed in the Real World because of Strategic Targeting by Candidates

independents, and twenty Republicans. Because the Democratic and Republican candidates are acting strategically, both candidates target the same number of advertisements for independents, but only the Democratic candidate airs ads that are seen by the Democratic voters and only the Republican candidate shows ads to the Republican voters. Imagine that before the election began, every Democratic voter already had a .8 probability (or 80% chance) of voting for the Democratic candidate and every Republican voter had a .8 probability of voting for the Republican. Independent voters were equally likely to vote for either candidate (thus, they had a probability of .5 of voting for the Democrat and a probability of .5 of voting for the Republican).

You might see the problem that this will create for drawing a causal inference about the effect of advertising on vote choice. If we just look at the relationship between these two factors while ignoring partisan identification, we would erroneously observe a strong effect for advertising. Even if exposure to advertising had no effect at all, we would find that when a voter saw only Democratic advertisements, their probability of voting for the Democrat was .8; when they saw only Republican advertisements, their probability of voting for the Republican was .8, and when they saw advertisements from both parties, their probability of voting for either candidate was .5. Based on these data, we might conclude that when a voter sees only one candidate's advertisements, her probability of voting for that candidate increases by .3 (from .5 to .8). But we would be wrong: the advertising did not change anyone's vote preferences; the effect was merely **spurious**.

The problem with the observational study is that we had no control over which voters saw which advertisements. A voter's party identification, which

influences that individual's vote choice, was also closely related to which candidate's advertisements she saw. Unless we accounted for voters' party affiliations

EXPERIMENTS

in some way, we would be misled into thinking that advertising had an effect when it did not. Of course, there are ways to account for such confounding factors in observational studies, and those will be discussed in the following chapters. But the additional problem is that the world we observe is rarely so simple and it may often be the case that we are unable to account for all of the confounding explanations. Indeed, we often do not even know what all of those confounding factors are.

By providing researchers with control over the data-generating process, experiments offer a way to remove any confounding factors rather than attempting to account for all of them. To see why this is the case, let's return to the example above. Imagine that at the beginning of the campaign both candidates decided that they wanted to learn whether it was worthwhile to spend so much money on advertising. Thus, for this campaign they decided to give control over their advertising budgets to an enterprising researcher. With control over the data-generating process, the researcher now has to decide the best way to allocate the advertisements so that it is unlikely that there will be any confounding explanations for any vote differences he observes. If the researcher makes the allocations randomly, then, by definition, they will be uncorrelated with other factors. The researcher should now be free of any worry about alternative explanations if any differences are uncovered.

Figure 5.3 shows how this could work in practice. The same sixty individuals have been divided into the same three different advertising conditions: one group will receive the Democratic advertisements, another group will receive the Republican advertisements, and the last will receive both ad-

vertisements. Now, however, the three groups have been randomly assigned rather than selected by the candidates. Because of this, there will always be about the same percentage of Democrats, Republicans, and independents in each condition (as long as the sample size is sufficiently large).⁷ As in the previous example, half of the sixty individuals will vote for the Democrat and half will favor the Republican, and because of randomization, each group should also be divided 50/50. If advertising has no effect at all on vote preferences, then we should observe relatively similar support across each condition, or across each of the three groups. However, if advertising is influential, then the group receiving the Republican advertisements should have a higher percentage of Republican voters and the group receiving the Democratic message, a higher percentage of Democratic voters. Thus, by allowing the researcher to control the distribution of advertising rather than the candidates, and by that researcher randomly assigning which individuals saw which advertisements, we can more clearly determine whether campaign ads are effective (or not).

⁷ The point about a large enough sample size (or subject pool) is an important one that we address below. For the time being, it is sufficient to say that, with a small sample size, one group might have a disproportionate amount of one party just by random chance.

EXPERIMENTS

D	R	D	I	I	I	R	D	I	I
I	R	R	I	R	D	D	R	R	D



Republican Ads

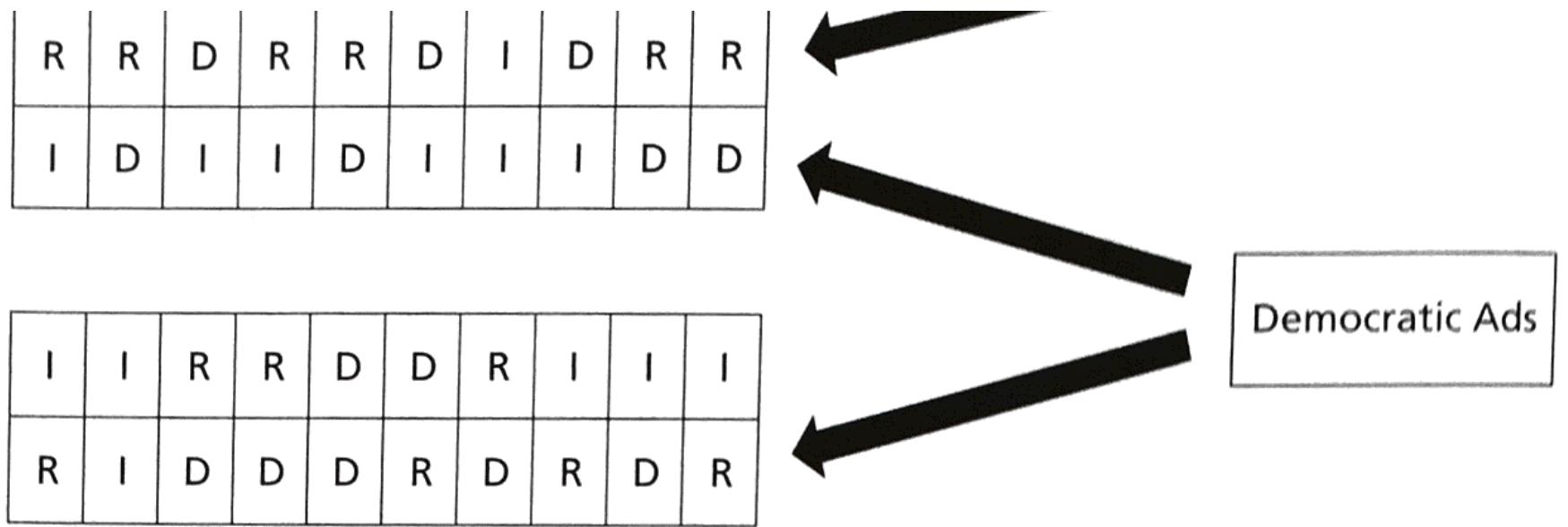


FIGURE 5.3 How a Researcher Might Distribute Advertising after Randomization

TYPES OF EXPERIMENTS

An experiment is defined by the ability of the researcher to exercise some control over the data-generating process. However, it is possible to execute an experiment in a variety of different ways. Generally speaking, political science research has been conducted via laboratory experiments, survey experiments, and field experiments. We discuss each in turn here.

Laboratory Experiments

The most commonly executed experimental design in political science has been the “**laboratory experiment**.” Rest assured that this is not as ominous as it may sound. Political scientists rarely (if ever) get involved in poking or prodding individuals with needles and we are aware of few political science experiments that have required the use of petri dishes. Rather, a typical laboratory that political scientists operate with tends to look like nothing more than a sophisticated computer lab. On those computers, political scientists run pro-

grams that ask participants to engage in a wide variety of tasks designed to mimic political situations.

While political scientists were slower to adopt laboratory experiments compared with disciplines like economics and psychology, they have now been used to focus on a wide array of questions using a variety of approaches. A relatively simple form of a laboratory experiment might invite subjects to watch television advertisements for some fictional candidate(s). Subjects would be randomly assigned to view different commercials, and, typically, the researcher would assign one group (called the **control group**) to see no advertisements

EXPERIMENTS

Box 5.1: Genes Affect Your Political Attitudes: A Different Kind of Experimental Design

While experimental methods are relatively new to political scientists, an even newer strand of research is focused on linking our understanding of politics to biology and genetics. One of the first major studies in this vein was authored by John Alford, Carolyn Funk, and John Hibbing, and published in 2005.² The authors asked a simple question: “Are political orientations genetically transmitted?” Previously, political scientists had not seriously considered a role for genetics, instead focusing on understanding how childhood socialization played a role in attitudes taken by citizens later in life.

To examine whether genetics played any role in shaping future attitudes, the authors used what can best be described as a natural experiment. To do this, they compared monozygotic (or “identical”) twins with dizygotic (or “fraternal”) twins. The key to this comparison is that monozygotic twins share 100 percent of their genetic material while dizygotic twins share only 50 percent (on average). To determine how important genetics were in causing political attitudes, the authors examined the rate at which identical twins shared political views to the rate at which fraternal twins did so. They found that identical twins were much more likely to share opinions on a range of politi-

cal issues, a finding they attributed to the fact that identical twins shared a much higher percentage of genetic material compared with fraternal twins. In fact, the authors found that genes play a substantial role in accounting for issue attitudes. They conclude, "We find that political attitudes are influenced much more heavily by genetics than by parental socialization . . . genetics accounts for approximately half of the variance in ideology, while shared environment, including parental influence only accounts for 11%."^b

Of course, no methodological approach is perfect, and the twin design that this study was based on has received some criticism.^c One basic critique with twin studies is that they rely on an assumption that fraternal and identical twins are raised in the same types of environments (the Equal Environment Assumption). But research has found that identical twins are treated by parents and others as if they are more alike compared with fraternal twins. Thus, it may be difficult to disentangle just how much of the increased similarity between the attitudes of identical twins is due to the fact that they are treated differently rather than the fact that they share more genetic material.

- . John R. Alford, Carolyn L. Funk, and John R. Hibbing, "Are political orientations genetically transmitted?" *American Political Science Review* 99 (2005): 153–167.
- . Alford, Funk, and Hibbing, "Are political orientations genetically transmitted," p. 164.
- . Evan Charney, "Genes and ideologies," *Perspectives on Politics* 6 (2008): 299–319.

I. After watching the advertisements (or not watching them in the case of control group), subjects would be asked, for example, to evaluate the date(s). Because the groups were randomly assigned and because they were identical in every way but the advertisement, if the researcher detected

differences between the groups who saw advertisements (called the **treatment groups**) and those in the control group, then she could conclude that those differences had been caused by the advertising.

Stephen Ansolabehere and Shanto Iyengar used a series of laboratory experiments to examine the extent to which citizens were influenced by negative advertising.⁸ Subjects were recruited to watch a local television news program for fifteen minutes. During a commercial break for that program, subjects would see an advertisement for a candidate who was currently running either for president or statewide office. Depending on which condition subjects were randomly assigned to, the advertisement they saw was either positive or negative in tone. This was the only aspect of the videos shown to subjects that differed across conditions. Ultimately, one of the key findings from this study had less to do with persuasion than it did with de-mobilization; specifically, the authors found that when subjects viewed negative advertising, they became less likely to want to vote at all.

The above example is a relatively simple experiment because individuals are examined as unitary actors. That is, a subject comes into the laboratory, is subjected to some treatment, asked some questions, and is then dismissed. More complicated laboratory experiments have studied decision-making processes by having multiple subjects interact with each other in the laboratory during the experiment. Often, these experiments are constructed to examine the extent to which individuals are able to reach collective decisions under different rules and circumstances. Consider for example, the research question posed by Nobel Laureate Elinor Ostrom and her colleagues: Under what conditions can members of a community overcome the obstacles to collective action and self-govern common pool resources? If you'll remember the discussion from Chapter 1, the tragedy of the commons posited that when individuals derive private gains from using the commons but only share the costs with all other

users, they will overexploit the commons. This poses a problem of collective action, as the actions of one individual depend on the actions of another. If I live in a community where everyone is exploiting the resource as fast as possible, then I had better try to get my share before it is gone. If, however, I live in a community where other users sustainably use the resource and levy sanctions against those who over-exploit, then I might adapt a different strategy.

Using both small- and large-n studies, Ostrom and her colleagues found that self-governance can and does occur. (We will take a look at some of these studies in the next chapter.) The next question was under what conditions does self-governance occur? This was a harder research question to answer. The world is a messy place and determining *why* someone decided to limit his share of a resource is not an easy question to answer. The controlled environment of the laboratory was, therefore, particularly attractive. In their article “Covenants with and without a sword: Self-governance is possible” Ostrom

EXPERIMENTS

and her colleagues James Walker and Roy Gardner specifically wanted to test the impact of (1) communication and (2) sanctions, or *covenants* and *swords* in the terminology of Thomas Hobbes.⁹ For Hobbes and many social scientists, covenants not backed by the credible threat of sanctions are meaningless. As he writes, “And Covenants, without the sword, are but words, and of no strength to secure a man at all.”

To test these two factors, Ostrom and her colleagues developed an experimental game that simulated the commons. The rules of the game ensured that the research participants were interdependent on one another. Game participants, undergraduate economics students, could either invest in maintaining the commons or they could invest in some alternative activity. If they invested

in the alternative activity, they were sure to get a return. If they invested in the commons, their return would depend on how much others invested in the commons. If other game participants also invested, they would do very well, but if other game participants did not invest, then they would be better off investing in the alternative activity. The game tried to simulate reality in that participants invested actual money and got to keep money won through the game. The experiments had one control group, where participants could not communicate and could not sanction, and three treatment groups: one that could only communicate, one that could only sanction, and one that could both communicate and sanction.¹⁰ After repeated rounds of play, game participants in the control group only earned 32 percent of what they potentially could have. They performed on par with the group that could only sanction, which earned 38.8 percent of the potential total. The group that could only communicate did surprisingly well, even when the stakes were high, and earned 75 percent of the total. The group that could sanction and communicate earned 97 percent of the total. The authors were not surprised to find that the sanction and communication group did well, but their findings that communication was a more important factor than sanctions contradicted a good deal of political theory that had posited the opposite.

Ostrom and her colleagues were able to confidently identify the importance of communication because the laboratory is a relatively controlled venue. This control provides the experiment with a great deal of **internal validity**. In the laboratory, each subject tends to operate in the same environment with the only differences being those that the researcher controls. However, laboratory experiments tend to be criticized for their lack of **external validity**. This critique is largely based on the fact that, for a variety of reasons, subjects may respond quite differently to treatments in a controlled venue like the laboratory

- ✓ Elinor Ostrom, James Walker, and Roy Garner, "Covenants with and without a sword: Self-governance is possible," *American Political Science Review* 86 (1992): 404–417.
- 10 In the actual experiment, the authors varied additional factors, including the total amount of money at stake, how much the participants were able to communicate, the number of rounds of play, and the option to choose to have a sanctioning device for a small fee.

EXPERIMENTS

compared with out in the real world. Take the television advertising experiment for example. When subjects are invited to come to a laboratory to watch campaign advertisements, they likely will watch those ads much differently than they would in their own homes. In the laboratory, subjects know that they are being watched, so they may pay more attention to the advertisement than they would at home, where their attention would be divided by children, pets, etc. Indeed, how many people (who aren't political science majors) do you know who watch campaign ads intently?

In order to isolate a cause and effect, laboratory experiments are designed to simplify the world, but the real world is messy. Individuals in the real world are also bombarded with emails and phone calls from the candidates; they read newspaper articles about the candidates; and they have conversations with friends and family about the candidates. All of these things would likely have some effect on their vote choices and the competing sources of information would likely mute the effects of advertising in the "real world."

Political scientists who conduct laboratory experiments often address the external validity critique in several ways. First, experimental political scientists often note that, while their experiments may be somewhat lacking in external validity, the gains made on internal validity are worth it. Additionally, there is great value in approaching political questions with a variety of techniques that have different strengths and weaknesses. Laboratory experiments can help us uncover causal mechanisms and generate stronger causal inferences while

observational studies can help us determine how generalizable those inferences might be in the “real world.”

Second, it is often possible to design laboratory experiments so that they come closer to simulating reality and political scientists are becoming increasingly adept at doing this. For example, Richard Lau and David Redlawsk created a computer program that helped to better account for the variety of factors competing for a voter’s attention during a campaign. As the authors explain:

We have designed an interactive experimental paradigm to study voter decision making that captures the crucial features of modern political campaigns: They are media-based; they provide an overwhelming amount of relevant information, some of which voters choose to expose themselves to, some of which comes to voters without any conscious decision to learn it, and much of which is simply missed; and they are dynamic, in the sense that information available today may be gone tomorrow.¹¹

Subjects in these experiments sit at a computer screen while links to items about the candidates scroll down the screen. Subjects can click on those links

11 Richard R. Lau and David P. Redlawsk, “Voting correctly,” *American Political Science Review* 91 (1997): 585–598, p. 587.

EXPERIMENTS

to read more, but, while they are reading about that information, other information is scrolling by in the background. Furthermore, at certain points during the experiment, a political advertisement appears on the screen and the individual can do nothing else but watch the ad until it is over. The idea

behind this elaborate set-up is that campaigns are full of information, and the better a researcher can approximate that environment in the laboratory, the more generalizable those conclusions will be for how voters behave in the real world.

Ansolabehere and Iyengar were also careful to design their experiments in order to maximize the external validity of their findings.¹² First, rather than show subjects advertisements from fictional candidates, they focused on real candidates and conducted their experiments during the actual campaign. Second, the researchers were careful to make the advertisements look realistic, often by making only minor adjustments to advertisements the campaigns were actually airing. Third, the advertisements were also presented to subjects in a setting that was as natural as possible. As the authors noted, “The viewing room was furnished with a couch, easy chairs, coffee table, and potted plants. Participants could snack on cookies and coffee while they watched the news, and in most cases participants came accompanied by a friend or co-worker.”¹³ Finally, the researchers were careful not to tell subjects about what they were studying until after the experiment was over. Instead, subjects were told that they were participating in a study about how individuals perceive the news. If subjects knew that they were participating in a study of television advertising effects, they may have paid more attention to the advertisements than would be typical. Indeed, in many political science experiments it is necessary to keep the goal of the experiment hidden from subjects until after they have completed the task.

A second threat to external validity in laboratory experiments is the nature of the subjects being studied in the experiments. Political scientists often rely on **convenience samples** for their studies, and most typically these convenience samples are made up of their own students. Students are in abundant supply on college campuses and researchers have discovered that many of them are more than willing to participate in research experiments in exchange for a small fee or extra credit. However, students also tend to differ from the

general population or American adults in several important ways. For example, college students tend to have more malleable opinions and attitudes on issues and greater cognitive capacity than the general population.¹⁴ As you

12 For more detail on how Ansolabehere and Iyengar address external validity, see pp. 20–22 of *Going negative*.

13 Ansolabehere and Iyengar, *Going negative*, p. 21.

14 David O. Sears, “College sophomores in the laboratory: Influences of a narrow data base on social psychology’s view of human nature,” *Journal of Personality and Social Psychology* 51 (1986): 515–530.

EXPERIMENTS

might imagine, these tendencies could influence the findings in many political studies. For example, political scientists often use laboratory experiments to examine whether framing an issue in different terms can cause citizens to take different positions on that issue. If, in fact, college students have weaker opinions on issues, then it may be easier to change their mind by framing an issue in a different way than it would be for a typical adult. Nevertheless, recent work by James Druckman and Cindy Kam suggests that the situations in which student subject pools pose a threat to external validity tend to be relatively limited, and that concerns about the use of this population in experiments may be over-stated.¹⁵

To increase the external validity of their studies, researchers often try to recruit non-student adult populations for their studies. Such an approach typically entails advertising the opportunity in local newspapers or websites and offering some non-trivial payment in exchange for participation. For example, for their experiments on negative advertising, Ansolabehere and Iyengar re-

cruited participants by advertising in local newspapers, distributing flyers in shopping malls, making announcements at offices and churches, and even telephoning individuals from voter registration lists. However, recruiting a sample in this way can be challenging for a variety of reasons. First, the costs associated with such recruitment efforts can be substantial. In order to get adults to participate in laboratory studies, the fee must be non-trivial. For example, Ansolabehere and Iyengar paid \$15 to each subject in 1992 (which would be over \$20 today), with the typical experiment lasting approximately one hour. Since they recruited approximately 3,000 subjects for their experiments, this amounted to a cost of \$45,000 just in payments to subjects. Advertising can also be costly, and there are substantial administrative costs involved in such an effort. Ansolabehere and Iyengar ran their experiments from two different three-room suite offices in different parts of Los Angeles. The researchers had to pay for the office space and for the furniture and other materials located in that space. They also had to pay individuals to run the experiments in each of those offices, with the experiments generally taking place between 10 a.m. to 8 p.m. every evening. And even with all this effort, the types of adults recruited in this way are still likely to be unrepresentative of the general population.¹⁶

15 James N. Druckman and Cindy D. Kam, “Students as experimental participants: In defense of the ‘narrow data base’,” in James N. Druckman, Donald P. Green, James H. Kuklinski, *et al.*, eds, *Cambridge handbook of experimental political science* (Cambridge, UK: Cambridge University Press, 2011), pp. 41–57.

16 In the Ansolabehere and Iyengar experiments, the individuals recruited were much more likely to have a college degree relative to the general population and they were also much more likely to be African American.

Survey Experiments

Survey experiments were initially conceived as a way of understanding what researchers mostly considered to be a nuisance for public opinion polling; namely, that those being polled tended to express different opinions depending on how a question was worded or what else had been asked earlier in the survey. For example, respondents asked about whether they approve of the job the president is doing in office tend to give somewhat different responses when the question is placed after an inquiry about economic conditions. Asking the question about the economy first tends to lead to lower approval ratings when the economy is doing poorly and higher ratings when the economy is doing well. Either way, the question about the economy is priming respondents to think more about the economy as they evaluate the president than they would have otherwise.

Pollsters used survey experiments to better understand some of these patterns. For example, they would randomly assign respondents into different conditions, where one-half would get the economy question first and the other half would get the approval question first. This would help a pollster to understand the nature and magnitude of the question order effects they were encountering, thereby making it possible to adjust for any bias. For a political scientist, the same type of experiment may yield important insights into how the public thinks about politics and how their opinions can be manipulated by elites.¹⁷

A question wording experiment can be used in a similar way. Take, for example, an experiment conducted by the Pew Research Center in 2002. The survey was conducted by telephone from August 14 to 25, 2002 with a na-

tional sample of 1,001 adults. At the time, the United States was debating whether to invade Iraq, and the survey asked respondents about that very issue. However, the pollsters at Pew were interested in understanding whether respondents would answer differently if given a differently worded question. One-half of the sample was randomly assigned to a version of the question that simply asked “Would you favor or oppose taking military action in Iraq to end Saddam Hussein’s rule?” The other half of the respondents received a question that asked, “Would you favor or oppose taking military action in Iraq to end Saddam Hussein’s rule, *even if it meant that U.S. forces might suffer thousands of casualties?*”

The differently worded questions led to very distinct responses from the public. When asked the shorter version of the question, which omitted the clause about the risk of casualties, 62 percent of respondents favored military

17 John Zaller and Stanley Feldman, “A simple theory of the survey response: Answering questions versus revealing preferences,” *American Journal of Political Science* 36 (1992): 579–616.

EXPERIMENTS

action while 24 percent opposed it (14 percent were undecided). However, when the question wording included the phrase about the risk of casualties, support for military action dropped to 43 percent, with 42 percent opposed and 16 percent undecided. The different results produced by variations in question wording was important in its own right, particularly as it underscored how tenuous public support for an Iraqi invasion really was. However, the experiment had additional value to political scientists by demonstrating that Americans often fail to consider the risks inherent in policy actions unless specifically primed to do so.¹⁸

While survey experiments were a relatively uncommon research approach in much of twentieth-century political science, they are quickly becoming more widely used by political scientists for two main reasons. First, surveys are becoming an increasingly affordable way for political scientists to collect data. In the past, most survey data was collected by human interviewers either in person or by telephone. While many polls are still conducted in this manner, political scientists are increasingly turning to internet surveys. Such surveys can generally be conducted at a fraction of the cost of traditional polls because it is not necessary to employ individuals to ask respondents questions.¹⁹ Second, the use of internet experiments also expands the types of treatments that political scientists can apply to respondents. With surveys conducted over the telephone, researchers were largely limited to changing the wording of a question and observing whether the different wording influenced the types of answers respondents provided. Internet surveys allow for a more dynamic and flexible survey where respondents can be shown different pictures or videos during the interview. Thus, to study advertising effects, a researcher could embed video of an advertisement during a survey and ask questions about the candidates featured in the ad following that video.²⁰

Even with modern internet survey technology, survey experiments are typically more limited than laboratory experiments with regard to the types of treatments that can be employed. For example, it is far more difficult to set up an experiment where subjects interact with each other to reach collective decisions with a survey compared to the laboratory. Yet, while survey experiments tend to be simpler in their construction, they have the potential to improve external validity. The increased external validity for these experiments comes from the types of subjects used; survey experiments are most commonly performed on a random (or representative) sample of American adults. Accordingly, researchers need not be as concerned about whether the

- 18 David L. Eckles and Brian F. Schaffner, "Priming risk: The accessibility of uncertainty in public policy decision making," *Journal of Insurance Issues* 34 (2011): 151.
- 19 Brian F. Schaffner, "Innovations in survey research," in Stephen K. Medvic, ed., *New directions in campaigns and elections* (New York: Routledge, 2011), pp. 39–58.
- 20 Lynn Vavreck, "The exaggerated effects of advertising on turnout: The dangers of self-reports," *Quarterly Journal of Political Science* 2 (2007): 325–343.

EXPERIMENTS

findings they uncover with their survey experiments are generalizable to the larger public.

While it is generally true that survey experiments are conducted on representative samples, experimental political scientists have begun conducting these types of experiments on convenience samples as well. With this approach, a researcher programs a survey using some web-based survey software to place a survey experiment at some URL. Some of the most common software programs include Survey Monkey (www.surveymonkey.com), Survey Gizmo (www.surveygizmo.com) and Zoomerang (www.zoomerang.com). The researcher then uses a variety of platforms to recruit people to take the survey. For example, an advertisement on Google or Facebook may invite people to participate. Even more effective is recruiting from a community of individuals registered at sites like Amazon.com's Mechanical Turk site. Mechanical Turk is a platform that connects individuals interested in completing small tasks for payment with people or companies seeking such assistance. Initially, the types of tasks that individuals performed on Mechanical Turk included tagging images to indicate what appeared on those photographs or looking at websites to determine whether the content is unsuitable for children. However, social scientists soon discovered that individuals on Mechanical Turk were also willing to participate in online surveys for a nominal fee.

A political scientist seeking to use the Mechanical Turk community to conduct an experiment would start by using an online survey program to develop a short questionnaire. Depending on the length of the survey, the researcher would then decide how much to offer individuals registered on Turk to complete the survey. For example, for a survey of approximately five minutes, the researcher might offer between 25 and 50 cents. The researcher can then set a limit on how many responses she is willing to accept and whether individuals will be allowed to complete the task more than once (for most experiments, researchers will want to limit participation to one time). Depending on the length of the task and how much is offered, researchers are often able to collect hundreds of responses within a week.

Despite the fact that over 100,000 Americans have registered to complete tasks on Mechanical Turk, the population of Turk users is hardly representative of the adult population. Turk workers tend to be younger, more educated, and are more likely to be female compared with the American adult population. There is also some evidence that Turksters are more likely to identify as Democrats and independents than they are to be Republicans. Nevertheless, in comparison with other convenience samples (like college sophomores), the Mechanical Turk population is much more representative of the population and, even more importantly, produces variation on variables like age, education, and interest in politics. Thus, for researchers interested in quickly gathering data from a survey experiment of a convenience sample of American adults, Mechanical Turk has become a viable option. Several studies have even found that a Turk sample tends to be more representative and attentive than

EXPERIMENTS

using college students as subjects.²¹ Furthermore, for the student researcher without access to an experimental laboratory, an online survey using the Turk

community may be an excellent option.”

Field Experiments

Until recently, **field experiments** had been used sparingly by political scientists. Yet, in many ways, they represent an ideal research design for demonstrating causal relationships with high levels of internal *and* external validity. A field experiment is an experiment that takes place out in the world, where the researcher lacks laboratory-like control over the environment. Individuals are generally selected to participate in such an experiment without their knowledge and often real-life political outcomes are observed. Perhaps the most famous recent examples of field experiments are those carried out by Don Green, Alan Gerber, and their colleagues.²³ Green and Gerber were interested in understanding whether candidates or parties could increase voter turnout by contacting voters and encouraging them to vote. Observational studies seeking to understand the effects of campaign contact on voter turnout suffered from some of the same issues mentioned above in the discussion of campaign advertising. Candidates and parties only target certain individuals for such appeals. Thus, Green and Gerber knew that they needed a way to randomize which voters were contacted and how they were contacted.

The Green and Gerber experiments have usually involved using lists of registered voters maintained by states or parties to randomly assign individuals into different conditions. They then cooperate with either political parties or interest groups to have some individuals receive an appeal to vote while others do not. After the election, they can return to the state’s vote records to determine whether, in fact, the individuals who received contact were more likely to vote than those in the control condition. Because such a large number of individuals have been randomly assigned into different conditions, if they find

differences between the treatment and control conditions, they can be confident that these differences can be attributed to the appeals. For the most part, Green and Gerber have found that face-to-face appeals are the most effective, while those sent by mail or by phone have only marginal effects.

21 Adam J. Berinsky, Gregory A. Huber, and Gabriel S. Lenz. “Evaluating online labor markets for experimental research: Amazon.com’s mechanical turk,” *Political Analysis* 20 (2012): 351–368.

22 Indeed, we have had much recent success advising undergraduate students who have used Mechanical Turk to conduct experiments for their research papers.

23 See, for example, Donald P. Green and Alan S. Gerber. *Get out the vote: How to increase voter turnout* (Washington, D.C.: Brookings Institution Press, 2008).

EXPERIMENTS

Green and Gerber have made a forceful case for field experiments as the “gold standard” of social science research:

Random assignment ensures unbiased inferences about cause and effect. Natural settings ensure that the results will tell us something about the real world, not just some contrived laboratory setting. Field experimentation would therefore seem to recommend itself as the most solid and unobjectionable form of social science and program evaluation.²⁴

Field experiments have been used to answer a wide array of research questions over the past several years. For example, one study randomly assigned individuals in Virginia to receive free subscriptions to either a conservative or

liberal newspaper during a campaign to examine whether newspaper coverage influences how individuals vote.²⁵ In another field experiment, a researcher convinced candidates in Benin to make different types of appeals during a campaign (narrow versus broad appeals) to determine whether these different campaign messages led to different results.²⁶ Another set of researchers conducted a field experiment in Rwanda to examine whether using radio-based appeals to encourage citizens to be less deferential to authorities led those citizens to express more dissent with the government.²⁷

A recent study even used a field experiment to address the influence of campaign advertisements on vote choice.²⁸ In 2006, Texas Governor Rick Perry (R) agreed to allow \$2 million worth of his television and radio advertising decisions to be randomly assigned by political scientists. This control was ceded to researchers during a three-week period in the primary phase of his re-election campaign in eighteen of the twenty television markets in the state of Texas.²⁹ The political scientists involved in the study randomly assigned each market into different conditions so that decisions about when Perry's advertising would begin in that market (and how much advertising would appear) would be uncorrelated with other factors. They then conducted daily surveys

24 Green and Gerber, "The underprovision of experiments in political science," p. 94.

25 Alan S. Gerber, Dean Karlan, and Daniel Bergan, "Does the media matter? A field experiment measuring the effect of newspapers on voting behavior and political opinions," *American Economic Journal: Applied Economics* 1 (2009): 35–52.

26 Leonard Wantchekon, "Clientelism and voting behavior: Evidence from a field experiment in Benin," *World Politics* 55 (2003): 399–422.

27 Elizabeth Levy Paluck and Donald P. Green, "Deference, dissent, and dispute resolution: An experimental intervention using mass media to change norms and behavior in Rwanda," *American Political Science Review* 103 (2009): 622.

28 Alan S. Gerber, James G. Gimpel, Donald P. Green, *et al.*, "How large and long-lasting are the persuasive effects of televised campaign ads? Results from a randomized field experiment," *American Political Science Review* 105 (2011): 135–150.

29 The campaign did not include the control of Houston or Dallas, Fort Worth, the two largest

the campaign did not want to cease control of Houston or Dallas–Fort Worth, the two largest markets in the state.

EXPERIMENTS

across the state to see how support for Perry fluctuated across each market. The authors did find that advertising had strong, but short-lived, effects on support for Perry. When advertising was introduced in a market, support for Perry increased, but that increase decayed after a few weeks.

It is not often that candidates are willing to turn over their advertising decisions to political scientists, but the opportunity was an excellent one for expanding what we know about advertising effects during campaigns. As the authors note:

This research is not the first study to show advertising effects, but demonstrating these effects with a field experiment is an important advance because the research design sidesteps criticisms that are often levied against other research methods. The large effects that are observed in the laboratory are routinely challenged on the grounds that they fail to tell us how media exposure translates into votes in the context of an actual campaign. . . . The large effects found in observational studies are similarly open to the charge that campaigns target their ads strategically.³⁰

Because this experiment was carried out during an actual campaign, it helps to bolster the findings from earlier laboratory experiments as well as those produced from observational studies.

Of course, field experiments can be difficult and often very costly to conduct. For example, it costs a lot of money to randomly send campaign mailings or some other information to hundreds (or thousands) of citizens. It can also cost a great deal of money to collect information on the dependent variable, particularly if doing so involves conducting a survey. Thus, for the aver-

age student, a field experiment is probably only viable when conducted in cooperation with either (1) a professor with a substantial research account or (2) an organization that would like to evaluate the effectiveness of its activities. Based on our informal assessments of most of our colleagues' research accounts, the latter situation is the one most likely to apply to the vast majority of students. But this possibility should be taken seriously by the enterprising student researcher. Think about groups (or politicians) for which you have an internship. The people at these organizations may very well be interested in a project that would help them understand how to persuade more citizens to support their cause. If you can design a study that would accomplish that goal, and you can explain how it would benefit their organization (as well as your own research goals), then they may very well allow you to have some control over their activities.

30 Gerber *et al.*, "How large and long-lasting are the persuasive effects of televised campaign ads?" pp. 147–148.

EXPERIMENTS

Box 5.2: Are Field Experiments Ethical?

While field experiments hold enormous potential for their ability to capitalize on high levels of internal and external validity, some scholars have questioned to what extent they are ethical. One

ethical concern has to do with the fact that, in most field experiments, the subjects do not know they are participants in a study. This is not the case in laboratory or survey experiments, where individuals can opt out of a study (no individual can be forced to participate in an experiment or a survey). In field experiments, it is difficult, if not impossible, to inform subjects that they are participating in a study in the first place. Imagine researchers trying to contact every individual in the state of Texas before they conducted the advertising experiments there in 2006. Of course, if a field experiment has little chance of causing any harm to an individual, this inability to opt out may not be of concern. But what of an experiment like that conducted by Paluck and Green in post-genocide Rwanda? In that experiment, some communities were assigned to hear radio programs that would encourage people to be less deferential to authority and the researchers found that listening to these programs did make individuals more willing to express dissent. But what if these expressions of dissent led to some individuals being arrested or even just shunned by the community? Such an outcome could have significant social and economic consequences for those individuals.

A second (somewhat related) ethical concern is that a field experiment may hold the potential of influencing a social outcome, such as an election result. For example, imagine a field experiment designed to see whether voters could be mobilized by particular messages. If the mobilization was successful, and it succeeded mostly by mobilizing low-income voters who were more likely to vote Democratic, then a close election could be influenced by that effort. Then again, most field experiments generally produce rather small effects and, since they are randomized, they should not have a clear bias in any given direction. Furthermore, it is generally the case that scholars are cooperating with groups that would have made those same efforts anyway; the presence of the scholar in these situations is merely to persuade the organization to include some degree of randomization in the distribution of those resources. Nevertheless, scholars should make themselves aware of the potential ethical questions surrounding field experiments, and endeavor to produce experiments that are as ethical as possible. Following the proper Human Subjects Review process at your university will help to ensure that your experiments are as ethical as is warranted.

A few examples of successful field experiments executed at little (if any) cost by the researchers may help to illustrate this point. Daniel Bergan conducted a study of lobbying by convincing a coalition of interest groups in New Hampshire to randomly determine which state legislators would receive grassroots

email appeals from their members and which would not.³¹ The appeals were managed by the interest group and Bergan only needed to download the final

31 Daniel E. Bergan, "Does grassroots lobbying work? A field experiment measuring the effects of an e-mail lobbying campaign on legislative behavior," *American Politics Research* 37 (2009): 327–352.

EXPERIMENTS

roll call vote from the legislature's website to see whether the lobbying had an effect on how legislators voted on the relevant bill (it did). In another field experiment, Allison Dale and Aaron Strauss cooperated with interest groups to randomly determine whether individuals on their mobilization lists would get a text message reminder about voting and what kind of message they would receive.³² The text message campaign was carried out by the organizations and the researchers needed only to work with a voter file firm to determine the percentage of individuals in each condition that actually voted. Even the \$2 million television advertising experiment described above would have been utterly impossible for political scientists to carry out on their own, without the cooperation of one of the candidates. What these and other recent field experiments demonstrate is that, with a lot of initiative and a little luck, it is possible to conduct an excellent field experiment without breaking the bank.

■ DESIGNING THE EXPERIMENT

The design stage is make-or-break for an experiment. Unlike the analysis of

observational data, which is a more dynamic process where the researcher can often add new variables or conduct more interviews to account for mistakes committed in the early going, a poorly conceived experiment is typically impossible to salvage. If something goes awry with an experiment, then there is usually no other choice but to try again.

On its face, an experiment may seem relatively straightforward to design. After formulating hypotheses, the researcher should be able to easily identify which independent variables need to be manipulated in the experiment. For example, if the researcher hypothesizes that seeing more of a candidate's advertisements will make an individual more likely to vote for that candidate, then exposure to advertising is the variable that needs to be manipulated in the experiment. Once this variable has been identified, then the researcher must focus on how much variation needs to be introduced on that variable. The researcher might decide that two conditions are sufficient—in one condition subjects would see one advertisement from the candidate and in the other condition subjects would not see any advertisements. This is the most simple experiment that a researcher could conduct, as it is limited to just two conditions. Of course, the researcher may decide that it is important to see how different amounts of advertising may have different effects. This can be tested by adding more conditions to the experiment; one additional condition may have a group of subjects see five advertisements from the candidate while another group might see ten.

To conduct a simple experiment like the one just described, the researcher would need to begin by developing the treatment (in this case, an advertisement

32 Allison Dale and Aaron Strauss, "Don't forget to vote: Text message reminders as a mobilization tool," *American Journal of Political Science* 53 (2009): 787–804.

for a hypothetical candidate). The next step would be recruiting subjects to participate in the experiment. As noted above, this could involve soliciting participation from students on a college campus or perhaps recruiting participants online at a site like Amazon's Mechanical Turk. The number of participants that a researcher must recruit for an experiment depends on a number of factors, including the type of statistics that will be compared (means or proportions), the size of the differences the researcher expects to find, and the number of conditions the experiment has. While there are formulas to help provide guidance on these matters, a good rule of thumb is to make sure that you have at least thirty to fifty subjects in each condition. Thus, for the hypothetical advertising experiment that includes four conditions, the researcher would want to try to recruit at least 200 subjects.

Once the subjects have been recruited for an experiment, the researcher must assign them to one of the conditions. As explained above, it is important that this assignment be done randomly to help ensure that any differences observed across the groups are attributable to the treatments imposed by the researcher rather than some other factor. The researcher can use a random number generator available online or through programs like Excel, Stata, SPSS, or SAS; the randomization could also be as simple as drawing numbers from a hat, flipping coins, or rolling dice.

It is worth noting that randomization does not always work as well in small samples, so, when subject pools are smaller, the researcher might pursue one of two different strategies. First, the researcher could do the randomization as planned and check to make sure the groups are roughly equivalent on the types of variables that might be correlated with the dependent variable. For example, in the advertising experiment, the researcher might want to make sure that each condition has a roughly similar percentage of Democrats, Republicans, and independents. Ideally, the researcher would discover significant unbalance before the experiment is executed; however, if this is not possible, then knowing the nature of the unbalance will at least allow the researcher to

control for these rival explanations when analyzing the results.

A second possibility is to conduct what is called block randomization. Block randomization is a way of making sure that your groups will be balanced on some important variable (or variables). For example, if you are concerned about having a lack of balance on party identification, then you could start by taking four Democrats from your subject pool and randomly send one of these four Democrats to each of the four conditions. Then repeat this process until you have sent all the Democrats to one of the four conditions and follow by repeating the same process for Republicans and independents. Proceeding in this fashion will guarantee that your groups will be relatively even on that variable.

Once subjects have been assigned to a condition, they should be administered the appropriate treatment (or lack thereof if they are in the control group). Following the treatment, it is necessary to measure the dependent variable of interest in some way. In the field experiments conducted by Green and Gerber, this was done simply by obtaining the official voter files from the

EXPERIMENTS

state and then looking to see who voted and who did not vote. In a laboratory or survey experiment, the dependent variable is typically measured by asking questions of the subjects following the treatment. For example, after showing subjects a candidate's advertisements, you might ask how likely they would be to vote for the candidate who aired the advertisement. Subjects could place themselves on a scale ranging from 0 (not at all likely) to 10 (very likely). Additional questions could ask subjects how likeable and competent the candidate seemed (perhaps using similar scales). Once you have recruited a subject to take the time to participate, it generally makes sense to ask several questions of him after the treatment since you may find that the treatment influences some types of attitudes or opinions, but not others. The addition of such ques-

tions can also be useful in hiding the subject of the experiment.

ANALYZING AND PRESENTING RESULTS FROM AN EXPERIMENT

While experiments require a great deal of thought and planning to execute, they have the benefit of producing results that are relatively easy to analyze and present. The reason for this simplicity is that, by controlling the data generation process, the researcher has presumably removed the need to control for other variables or use complicated statistical techniques to establish a causal relationship. If the experiment was designed well, then to determine the effect of the independent variable on the dependent variable, one need only to compare values of the dependent variable across each of the conditions. The addition of a relatively simple statistical test will help to determine how confident you can be that these differences are more than just random fluctuations.

Once the experiment is completed, the researcher should generate summary statistics for each of the experimental groups. These summary statistics may be averages (if the variables are continuous) or proportions (if the variables are categorical). Once calculated, the statistics can be compared across groups to determine whether any differences exist. In these comparisons, the control group serves as the baseline, since this group did not receive any treatment at all. Then each treatment group can be compared with the control group as well as to each other.

Suppose we executed the experiment described above, where subjects were shown either no advertisements, one advertisement, five advertisements, or ten advertisements and then asked how likely they would be (on a scale of 0 to 100) to support the candidate featured in the ads. We would begin by taking each individual's score on the 0 to 100 scale and then averaging them to generate a mean value for each group. These mean scores are presented in Table 5.1.

The interpretation of these results would be rather straightforward. The average support for the candidate was only a few points higher for the group that saw just one advertisement, but support was about 15 points higher among subjects who viewed five advertisements.

EXPERIMENTS

TABLE 5.1 Average Scores in Hypothetical Advertising Experiment

	Control	One advertisement	Five advertisements	Ten advertisements
Average support	50.6	54.2	65.8	72.4

While a table like Table 5.1 is a perfectly reasonable way of presenting this information to someone reading about your findings, visual representations of results tend to be more impactful on readers. Thus, we might easily convert our findings from Table 5.1 into a figure like Figure 5.4. We chose a bar chart for this figure rather than a line graph because the line graph implies that we have data across the range of values on the x -axis. However, we really only know how subjects react to seeing either no, one, five, or ten advertisements, so a bar chart would be more appropriate here. We might have spaced the x -axis so that rather than each condition being next to each other there would be gaps between one and five, and five and ten, to account for the jump in the number of ads between those conditions. But the tradeoff of such an approach is that the bars would no longer be close enough to each other to easily gauge their relative sizes, so in this figure we kept the bars next to each other. Ultimately, this figure shows quite clearly that the biggest increase in support comes from moving from one advertisement to five advertisements. In other words, there are diminishing returns with increasing advertising; the first five

advertisements are much more influential than the next five.

To this point, we have merely examined the mean values for all subjects across conditions. However, there are often reasons to expect that the treatment effects might be larger for some groups than they are for others. For example,

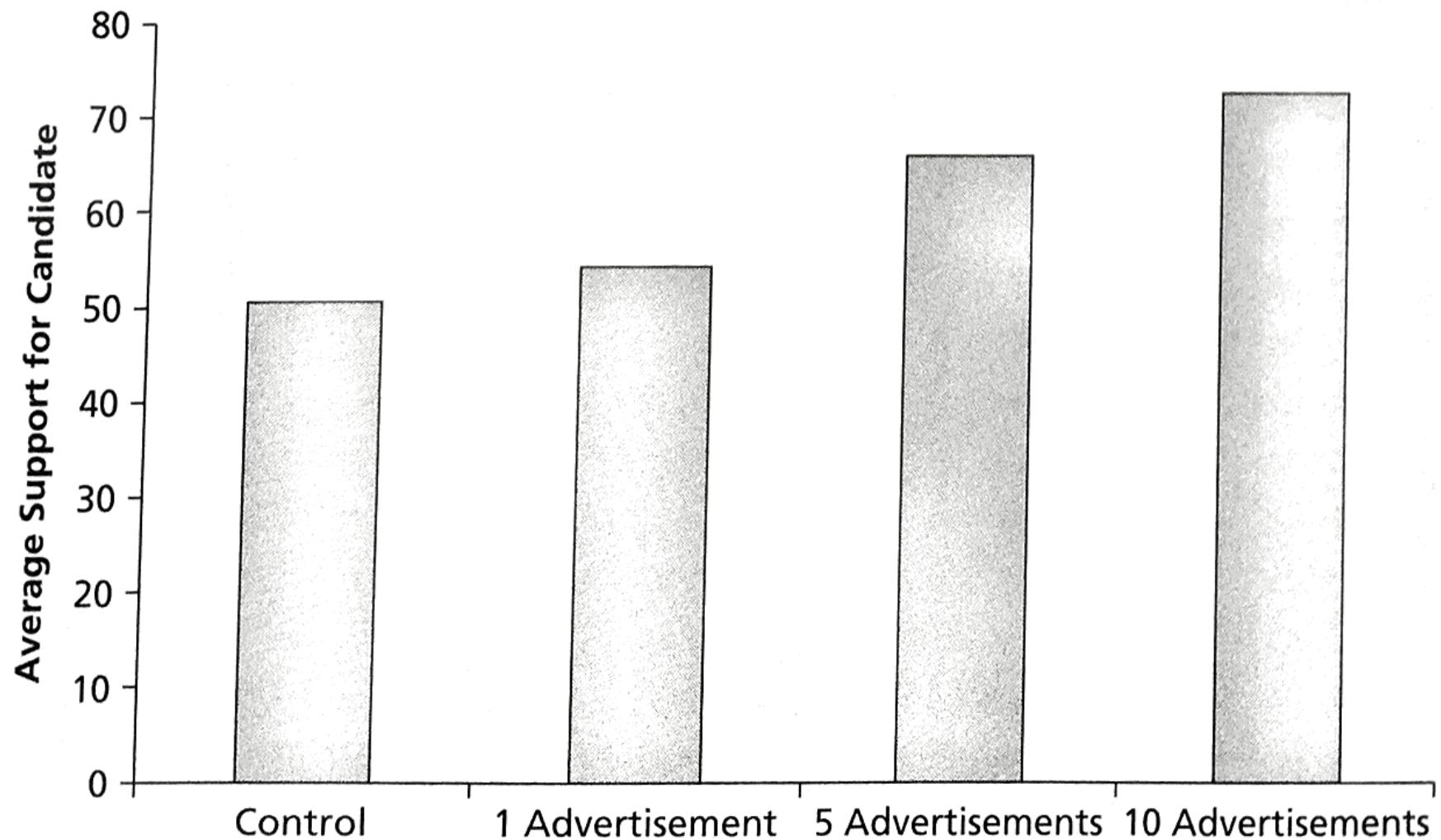
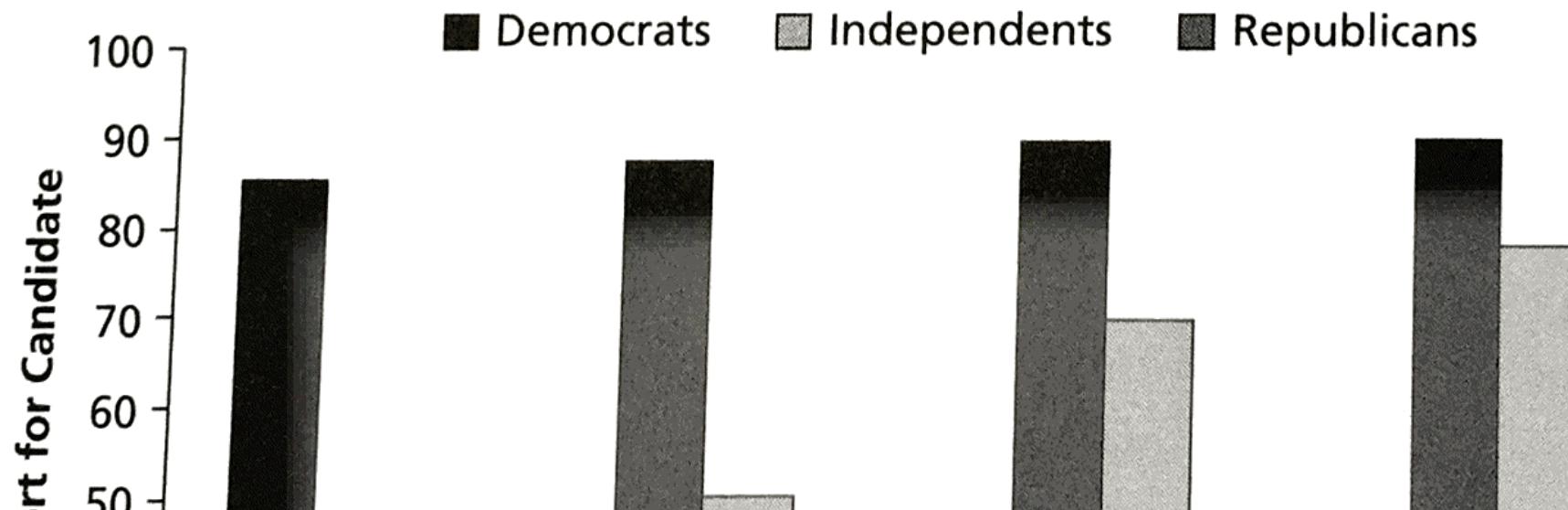


FIGURE 5.4 Example of the Presentation of Results from a Simple Experiment

people who pay less attention to politics may be easier to influence with political rhetoric; similarly, individuals who do not identify with either political party may be easier to persuade with advertising. Exploring the way that treatments affect some groups more than others is an excellent way to push the analysis beyond the simple reporting of differences across conditions and uncover additional results worthy of discussion. In many experiments, the treatment effects are small when comparing the effects among all respondents.

but, when subgroups are analyzed, larger and more significant differences may be revealed. For example, if independent voters are strongly influenced by advertisements, but neither Democrats nor Republicans are moved much, then looking at all subjects together could be a bit misleading. After all, the advertisements have a strong effect on independents, but independents might have only comprised about one-third of our subject pool.

Figure 5.5 presents how results for our advertising experiment might work differently among Democrats, Republicans, and independents who had viewed an advertisement from a Democratic candidate. Note that in this figure it is clear that advertising matters little for Democrats and Republicans. Democratic subjects already liked the Democratic candidate quite a bit, while Republicans already disliked her. More advertising from that candidate did little to change these predilections. Independents, on the other hand, were more influenced by the candidate's advertising, with their support being about 30 percentage points higher in the 10 advertisements condition compared with the control condition. The lesson here is that delving deeper into the findings to examine different treatment effects for different subgroups is often a rewarding exercise.



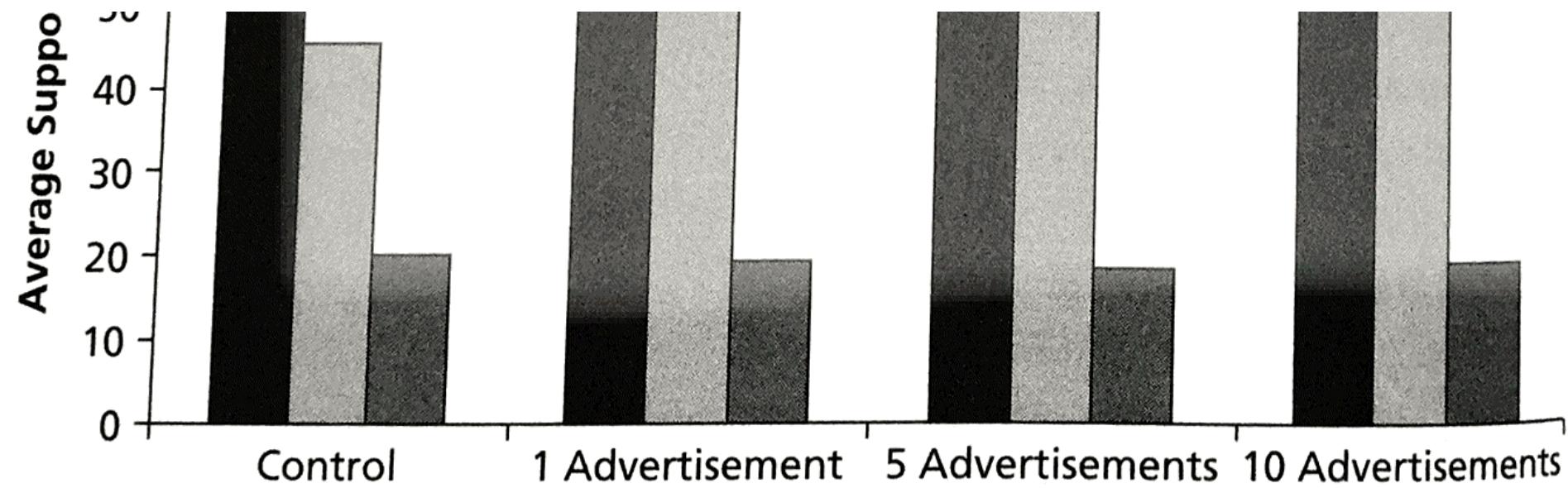


FIGURE 5.5 How the Effect of Advertising Might Differ by Partisanship in a Hypothetical Experiment

AVOIDING MISTAKES IN YOUR EXPERIMENT

There are many reasons why an experiment may fail. Recall that the strength of an experiment is in the creation of different groups that are essentially equivalent on everything other than the variable (or variables) being manipulated. To understand what effect the independent variable has on the dependent variable in your experiment, it is important to consider what the control condition should look like. Constructing a good control condition is often more difficult than it might initially sound. For example, what should a control group look like in a campaign advertising experiment? Should individuals assigned to the control group watch no advertisements at all? Or should they watch advertisements for some consumer product rather than the political advertisements? If you follow either of these strategies, then subjects in the control condition may be quite confused when you start asking them about candidates they have never heard of (since they never saw any of the relevant advertisements).

Political scientists Rebecca Morton and Kenneth Williams point out, “Many experiments do not have a clear baseline and in some cases it is not necessary.”³³ While this is certainly true in many situations, one should always take as much care as possible to create conditions that will produce as much variation in the key variable(s) as is warranted to draw useful conclusions. Thus, in an advertising experiment, it may in fact make sense to have a condition where some subjects see no advertisements at all. To deal with the problem mentioned above, all subjects could be asked to read an identical brief description of the candidates before watching ads. Thus, all subjects will start with the same baseline information about the candidates, and the only difference will be what advertisements the subjects see after reading that information.

A related issue that can arise in the design of an experiment is when more than one variable is allowed to vary across conditions. Consider a relatively simple experiment designed to determine whether female candidates are perceived differently than male candidates. A researcher might design an experiment where subjects read a brief description of a candidate for office and then evaluate the candidate along several dimensions. In one condition the candidate would be male, but in the other condition the candidate would be described as female. However, what if the researcher also designed the experiment so that the male candidate was a Republican and the female candidate was a Democrat. If the experiment was designed in this way, the variables for party and gender would be perfectly correlated and there would be no way to

³³ Rebecca B. Morton and Kenneth C. Williams, “Experimentation in political science,” in *Oxford handbook of political methodology* (Oxford: Oxford University Press, 2008), pp. 339–356.

differentiate their effects. Thus, the researcher would not be able to determine whether any differences in subjects' evaluations across conditions were attributable to the different gender of the candidate or the different party label.

While the example above is a relatively simple one that seems easy enough to avoid, it is remarkable how many experimental designs we have seen where this rule has been violated. This often happens when experiments become more complicated and there are many different pieces of information that have to be accounted for. At the same time, researchers are always cognizant of the limited number of subjects that they may be able to recruit for their experiments, so there is often a strong preference to create as few conditions as possible in order to avoid having too many conditions with too few subjects. Pre-testing is generally not useful for uncovering instances where too many variables vary across too few conditions. Instead, it is imperative on the researcher to think carefully about what pieces of information vary across conditions. It may even be useful to create a table or grid to ensure that every piece of information that changes has its own conditions.

Another common mistake made in the design of an experiment happens when the researcher develops a treatment that does not produce the variation in the independent variable that the researcher intended. Often this happens because the treatment is not strong enough and possibly goes unnoticed by the subjects. As an example, imagine an experiment that is designed to determine whether negative campaign advertisements make citizens less interested in politics. The researcher might design an advertisement that she thinks is clearly a negative attack on the candidate's opponent and then show this advertisement to subjects. But what if the experimental subjects do not perceive the advertisement in the way the researcher intended; perhaps they do not see the ad as negative at all. In this case, the experiment cannot be used to draw conclusions about the effects of negative ads since the subjects did not actually

perceive the ads as negative. In essence, the treatment is not operationalized in a way that really represents the theoretical concept the researcher is attempting to understand.

Researchers often take several different approaches to ensure that they do not create an experiment that fails to properly operationalize the key concepts. These strategies are best used in conjunction with one another for full effect. First, it is useful in an experiment to use a validity check to see whether subjects perceived the treatment in the way it was intended to be perceived. In the example of negative advertisements, respondents might be asked whether they thought the advertisement they just viewed was positive, negative, or neutral in tone. If most subjects correctly identified the advertisement as negative, then this would bolster the researcher's confidence in the results. Of course, if the researcher only does this when he runs the experiment, it will be too late if he discovers that most subjects do not see the advertisement as negative. Thus, a second strategy is to use a small subsample of subjects for a pre-test of the experiment. These subjects can participate in the experiment and then

EXPERIMENTS

Box 5.3: Studying Corruption with Experiments

The issue of corruption is an important one, and a wealth of political science and economic research has focused on understanding what mechanisms work for discouraging corruption. While many of these studies are large-n observational analyses, some scholars have used clever experimental approaches to gain a better understanding of this question. Benjamin Olken wanted to understand to what extent monitoring (with the threat of a sanction) could reduce the amount of corruption

observed in the building of public works projects.^a To do this, he cooperated with the Indonesian government to conduct a field experiment in conjunction with the government's allocation of block grants to construct roads in 608 different villages. Local officials receiving these grants may have an incentive to embezzle some of the funds allocated for the road projects. However, Olken assigned villages to different conditions. The control group was in a condition where the possibility of a government audit was normal (4 percent chance); villages in the experimental group had a 100 percent chance of being audited and were told this before they received their funds. Olken did find that less money was missing from projects in villages assigned to the treatment condition; however, more jobs were given to family members in these villages, suggesting that the nature of the corrupt acts simply changed.

a. Benjamin A. Olken, *Monitoring corruption: Evidence from a field experiment in Indonesia*, No. w11753 (Cambridge: National Bureau of Economic Research, 2005).

be fully debriefed by the researcher. During the debriefing, the researcher can ask these test subjects whether they viewed the treatments in the way that was intended and whether anything else about the experiment was confusing or otherwise problematic. Using the information gleaned from this debriefing, the researcher can then tweak the experiment before running it with the larger sample of subjects.

NATURAL EXPERIMENTS

To this point in the chapter, we have discussed experiments as research designs where the researcher has control over the data-generating process. Specifically, the researcher is able to manipulate the extent to which subjects are exposed (or not exposed) to some treatment. However, political scientists are also increasingly making use of what are often called **natural experiments**. Natural experiments differ from experiments in that the researcher actually does not

have control over assignment of the independent variable. However, good natural experiments can be analyzed as if they are experiments because some exogenous, approximately random process creates variation in the independent

EXPERIMENTS

variable in a way that is unlikely to be correlated with any other rival explanations. As Robinson, McNulty, and Krasno explain, “The hallmark of a natural experiment is a circumstance that creates some sort of arbitrary or random division of an observed population.”³⁴

To illustrate what makes a natural experiment different from both an experiment and a large-n observational study, we will use the example of a study conducted by Bob Erikson and Laura Stoker.³⁵ As you likely know, during the Vietnam War the United States drafted young men into military service. While the method of drafting individuals was initially idiosyncratic to local draft boards, in 1969 a standard national system was developed. This system was based on randomly assigning each day of the year a number between 1 and 366; individuals would receive the number that had been assigned to their birthday. Lower numbers were at high risk of being drafted while those with high numbers were at low risk of being drafted. Ultimately, individuals holding numbers 1 through 195 were called up in the draft.

Erikson and Stoker were interested in understanding how the risk of being drafted influenced the types of opinions young Americans developed towards the war and ultimately how it affected their vote choices and party identification in the long run. In a true experiment, the researchers would have to manipulate the variable of interest. Of course, it is quite unrealistic to expect researchers to be able to assign individuals to different conditions where they

have a higher or lower risk of being drafted, especially for a war that occurred several decades ago. Yet, even though the researchers did not have any control of the data generation process themselves, draft status was still determined by a random lottery and, therefore, should not be correlated with any other competing explanations. In essence, nature produced a quasi-experimental intervention that could then be exploited by the authors.

Erikson and Stoker took advantage of the fact that scholars had conducted a political socialization survey on young adults during the same time period, and they had re-interviewed these same individuals later in life. Because lottery status was randomly determined, Erikson and Stoker did not need to rely on sophisticated statistical techniques to determine whether draft status had an effect on political attitudes—they could simply compare the attitudes of those with lower draft numbers to those with higher draft numbers. Doing this, they demonstrated that “Males holding low lottery numbers became more antiwar, more liberal, and more Democratic in their voting compared to those whose high numbers protected them from the draft.”³⁶

- 34 Gregory Robinson, John E. McNulty, and Jonathan S. Krasno, “Observing the counterfactual? The search for political experiments in nature,” *Political Analysis* 17 (2009): 341–357, p. 349.
- 35 Robert S. Erikson and Laura Stoker, “Caught in the draft: The effects of Vietnam draft lottery status on political attitudes,” *American Political Science Review* 105 (2011): 221–237.
- 36 Erikson and Stoker, “Caught in the draft,” p. 221.

EXPERIMENTS

A strong natural experiment can be a very powerful approach to answering a research question. After all, natural experiments share many of the advantages of field experiments, particularly with regard to promoting high levels of

ages or new experiments, particularly with regard to promoting high levels of external validity. If the internal validity is also high, then a great deal can be learned. However, the key to utilizing a natural experiment is to ensure that the intervention is truly exogenous to the relationship being studied. In the case of a draft lottery, the exogeneity of the treatment is quite clear. An individual's risk of being drafted was assigned to him randomly, just as a researcher would have done it in a true experiment. But, in other natural experiments, the exogeneity of the treatment may not be quite as strong.

One dynamic that political scientists often capitalize on to construct natural experiments are how jurisdictional borders are drawn. For example, every ten years congressional redistricting means that many voters are moved from one congressional district to another. Stephen Ansolabehere, Jim Snyder, and Charles Stewart took advantage of this process to determine the extent to which members of Congress were able to build a personal vote for themselves.³⁷ They did this by comparing how well the incumbent performed in new parts of his district compared with those portions of his district that did not change. Of course, redistricting may not always be an entirely exogenous (or random) process. In many states, politicians themselves are responsible for drawing district lines and members themselves may have some influence on what their districts will look like over the next decade. Thus, in cases where the independent variable is not clearly exogenous and random, it is imperative on the researcher to make a case for why the intervention can be treated as if it is random and to demonstrate that the treatment is not correlated with other variables that could be alternative explanations for any patterns uncovered.

Let us return once more to the question of whether campaign advertising can persuade voters to support a candidate. We have already elaborated on how scholars have studied this question through various experimental designs, but Greg Huber and Kevin Arceneaux have shown that it is also possible to address the question using a natural experiment. Specifically, Huber and

Arceneaux took advantage of the fact that television advertisements cannot be targeted by campaigns in the same way as direct mail and phone calls. As the authors note, “Fortuitously, television broadcast signals, unlike campaign workers, have little regard for state boundaries. If a campaign purchases advertising in the Philadelphia media market to target voters in Pennsylvania, these broadcasts also appear on televisions in parts of Delaware and New Jersey.”³⁸

37 Stephen Ansolabehere, James M. Snyder, Jr, and Charles Stewart III, “Old voters, new voters, and the personal vote: Using redistricting to measure the incumbency advantage,” *American Journal of Political Science* 44 (2000): 17–34.

38 Gregory A. Huber and Kevin Arceneaux, “Identifying the persuasive effects of presidential advertising,” *American Journal of Political Science* 51 (2007): 957–977, p. 961.

EXPERIMENTS

Thus, if a candidate targeted advertisements to persuade voters in Pennsylvania, those advertisements would still be seen by citizens in Delaware, even though voters in Delaware were not the intended target of the ads.

The authors take advantage of this mismatch between market and state boundaries to examine whether advertising is persuasive. For example, Indiana was by no means a battleground state during the 2004 presidential election, but neighboring states like Michigan and Ohio were. Thus, when candidates aired advertisements in those states, some of those advertisements were inevitably seen by people living in Indiana near the borders of those states. Since exposure to these ads among people in Indiana was accidental, it was unlikely to have been correlated with other factors that would typically drive campaign decisions (such as pre-existing support for the candidates). Indeed, the authors

went the extra step of demonstrating that, in fact, the presence of advertising in these non-battleground states was entirely uncorrelated with previous support for either party. Ultimately, this allows them to discover that “advertising does a little to inform, next to nothing to mobilize, and a great deal to persuade potential voters.”³⁹

Another limitation with natural experiments is even more fundamental; by their very nature, natural experiments cannot be produced on demand. Furthermore, it is often difficult or even impossible to derive a set of hypotheses and then go about searching for a natural experiment that would be useful for testing them. Natural experiments simply do not exist for all (or even for most) research questions. Indeed, it is far more common for scholars to see random or approximately random interventions in the world and then think about what questions those interventions could be used to answer. Thus, when it comes to natural experiments, the research design is often executed before the research question is even posed. As we have often stressed in this book, the research process is very often not a linear one. Sometimes research questions are re-conceived to match the type of research design that is actually practicable, and sometimes research questions come directly from the discovery of excellent research designs. It is very possible that you may find a natural experiment that leads you to a research question you are interested in answering. The key strategy when looking for interesting natural experiments is to think about shocks that may have affected some portions of a population in different ways than others. Such shocks may be actually caused by nature—like extreme weather or earthquakes—or they may be induced by humans—like redistricting or draft lotteries. But as long as the shock is exogenous, it has the potential to be used as a natural experiment.

EXPERIMENTS**Box 5.4: The Nonlinear Research Process: How Experiments Often Make Us Rethink Our Theory**

As we stress throughout this book, it is often the case that, once we begin to analyze our empirical data, we are forced to rethink the theory we have developed or even refine our question. Experiments typically offer the best opportunities to do this, largely because of the strong causal inferences they allow us to make. Unlike with observational data where theory is crucial for directing us toward competing explanations we must account for, our confidence in experimental results is not dependent on a strong theoretical understanding of the process generating those results. For example, we might conduct an experiment where we heat a pot of water to 212 degrees Fahrenheit and find that it begins to boil when we do this. We do not need to have a theory about why the water boils at this temperature to be confident that it boils. Likewise, we need not have a theory of why advertising influences the vote preferences of individuals to be confident in experimental results that show that the relationship exists.

If experimental results do not produce the findings you expect, then it often makes sense to rethink the theory you developed to generate your hypotheses. For example, you may run an experiment and discover that advertising produces no increase in support for the candidate featured in the ads. Perhaps this finding indicates that people are impervious to advertising effects, or perhaps it makes sense to think more carefully about the nature of those effects. If you are confident that the experimental design is sound, then it is sometimes the theory that must be re-evaluated.

CONCLUSION

For maximizing internal validity, experiments offer the most ideal research approach. By controlling the data generation process, researchers can rule out competing explanations and ensure that any associations between the independent and dependent variables are the result of a causal relationship. Field experiments attempt to maintain the internal validity of a laboratory experiment while also increasing external validity by pulling the research out of the lab and into the real world. Of course, these also tend to be the most challenging types of experiments to execute.

It is often assumed by political scientists that there are substantial limits to the types of questions that experiments can be used to answer. Indeed, experiments have traditionally focused on addressing questions related to political psychology, such as how voters make decisions and the extent to which they are subject to elite manipulation through framing and persuasion. However, just because these are the types of questions that have been addressed by experiments thus far does not necessarily mean that they are the only questions amenable to experimentation. In fact, a little imagination can make it quite

EXPERIMENTS

possible for a researcher to apply experimental methods to a wide variety of questions. For example, recall that with field experiments the experimental subjects are unaware that they are participating in a study at all. Thus, it may be possible to conduct field experiments where politicians are the subjects, as Daniel Bergan did in his examination of lobbying effects in the New Hamp-

shire legislature. Additionally, it may be possible to use incentives to make it so that normal citizens act as if they were political elites, as Elinor Ostrom has often done in her decision-making experiments.

The point is that it is never wise to fully write off an experimental approach without giving the method considerable thought. At a minimum, we think it is a useful exercise for students to think about how they might analyze a question they are interested in using an experimental approach. The student may ultimately decide that such an approach is either inappropriate or unworkable for her particular question, but the mere process of thinking systematically about how an experimental approach might work will help the student think more carefully about how she should undertake an observational approach to answering the question at hand.

KEY TERMS

control group 122

convenience samples 127

data-generating process 118

external validity 125

field experiments 132

internal validity 125

laboratory experiment 122

natural experiments 143

spurious relationship 120

survey experiments 129

treatment groups 124

