Experiments and Causal Relations

2.1 Placing Experimental Research in Context

In typical discussions of estimating causality, social scientists who come from a statistics perspective often begin with a review of the experimental approach in an idealized setting that rarely exists, argue that the experimental approach as idealized is not feasible in social science, and then go on to discuss how causality is measured in observational data. For example, Winship and Morgan (1999) begin their otherwise excellent review of the literature in social science on measuring the effects of causes with the statement (p. 660), "sociologists, economists, and political scientists must rely on what is now known as observational data – data that have been generated by something other than a randomized experiment – typically surveys, censuses, or administrative records." This tendency to bracket off measuring causality in experimental social science from measuring causality in observational data presumes that experiments are "either-or propositions": a researcher can either conduct an "ideal" experiment, which we argue in this book would not be ideal for many questions in which political scientists are interested, or work with observational data.

Most of experimental social science is not the hypothesized ideal or classical experiment, usually with good reason. The bracketing off prevents a discussion of how causality is measured in experiments as they exist in social science and a realistic comparison of those methods to research with observational data. Moreover, many of the methods that are used to measure causality in observational data are also relevant for experimental work in social science, and researchers must understand the relationships between experimental design and these methods and how they interact.

As discussed in Section 1.5.1, if you ask political scientists what is the principal advantage of the experimental approach to political science, most

would answer that it can better measure causality than is possible with observational data. Yet, the relationship between the experimental approach to establishing causality and other methodological approaches to causality used in political science is not well understood and misunderstandings exist between experimentalists whose approach builds on work in social psychology or statistics and those whose approach builds on work in economics over how causality can be measured in experiments. A significant source of this lack of a common understanding is related to the welcoming-discipline nature of political science, as mentioned in Section 1.6.

Both the advantages and the disadvantages of being a welcoming discipline are also exhibited in the ways in which political scientists have addressed questions of causality. Political scientists rely on approaches to causality, which originated in statistics and biostatistics, sociology, psychology, and economics. This borrowing has benefited political scientists' research as the discussion of particular examples in Chapter 1 demonstrates. However, little has been written about how these approaches fit together or make sense for political science questions in a comprehensive sense (Zeng [2003] is exception) or how these approaches compare to the experimental approach. In this chapter, we explore how causality is measured (or not) in political science research and place the experimental approach within that context.

The importance of a discussion of measuring causality in a general sense for both experimental and observational data was highlighted by the recent exchange between Imai (2005) and Gerber and Green (2000) in the *American Political Science Review* on how to interpret the estimated effects in field experiments on mobilization. We thus discuss measuring causality in a general sense for a given data set, not making any assumptions about whether the data are observational or experimental. From this general perspective we then place the experimental approach to measuring causality – its advantages and disadvantages compared to observational data – in context.

2.2 Technical Background

The statistical and econometric literature on causality is built on existing knowledge in probability theory, procedures such as ordinary least squares, methods such as probit and logit, the maximum likelihood approach to data, nonparametric estimation techniques, and graph theory. As such, some of the material in this and following chapters necessarily refers to these techniques and for the more technical sections we assume that readers have prior exposure to probability theory and basic multiple regression techniques and

can work with simple mathematical models.[1] It is necessary to present this technical material because of the importance of the assumptions underlying the techniques for how we interpret data on causality, even assumptions that to an outside observer would appear to be innocuous. We attempt to make the material accessible to readers who have less exposure to these techniques in our interpretations.

Two important caveats about our discussion of causality are in order. First, we only discuss quantitative approaches to causality because our emphasis is on the measurement of causal relations through the analysis of data generated by manipulations either by an experimentalist or nature acting in a similar fashion. We do not address qualitative approaches to causality used in political science. Second, we primarily focus on studies of causality using cross-sectional and panel data rather than time-series data because most empirical studies with experimental data involve the use of such data. Specifically, when we have observations over a period of time, we are interested in cases for which it is reasonable to argue that it is more likely that the number of observations per time period approaches infinity faster than the number of time periods and panel data methods are appropriate.[2]

2.3 Causes of Effects Versus Effects of Causes

2.3.1 Causes of Effects and Theoretical Models

When addressing causality we must distinguish between investigations of the causes of effects and investigations of the effects of causes. If we are asking what causes turnout, we are asking a question about the causes of effects (turnout), but if we ask if making a voter more informed increases his or her probability of voting, then we are asking more narrowly about the effects on turnout of a cause (information). Heckman (2005a) presents a view from econometrics when he argues that ultimately we are interested in the causes of effects. He remarks (p. 2): "Science is all about constructing models of the causes of effects." Heckman also (2005b) contends that

"causality" is not a central issue in fields with well formulated models where it usually emerges as an automatic by-product and not as the main feature of a scientific investigation. Moreover, intuitive notions about causality have been dropped in

---

[1] In some discussions, knowledge of logit and probit is helpful, although not required.
[2] Hence, the asymptotic properties of estimators are evaluated as the number of observations approach infinity, holding time constant. In time-series data, the opposite is the case.

pursuit of a rigorous physical theory. As I note in my essay with Abbring (2007), Richard Feynman in his work on quantum electrodynamics allowed the future to cause the past in pursuit of a scientifically rigorous model even though it violated "common sense" causal principles. The less clearly developed is a field of inquiry, the more likely is it to rely on vague notions like causality rather than explicitly formulated models.

The emphasis on models of causes of effects as the primary goal of study is no doubt the main reason why Heckman advocates what he calls the structural approach to causality, which with observational data is close to the formal theory approach and which we explore in detail in Chapter 6.

In the formal theory approach to causality, an empirical researcher works with a model of the causes of effects from previous theoretical and empirical work and then evaluates that model (predictions and assumptions) with available data, either observational or experimental. The model usually makes a number of causal predictions rather than just one, but all are logically consistent with each other and with the model's assumptions. The causality in the model is often conditional to given situations; that is, some variables may be simultaneously determined. The evaluation of the model leads to further research, both theoretical and empirical. Sometimes theoretical investigators may think like Feynman; that is, envision situations that are beyond common sense in order to explore the logical implications of the model in these nonsensical worlds. Empirical investigations, however, tend to use applied versions of the model (although experiments can allow for the researcher to move beyond the observed world in the same way theory allows, if the researcher desires). This approach is also presented in political science by Morton (1999) and Cameron and Morton (2002) and is the basis of most laboratory experiments conducted by political economists and some by political psychologists (although with nonformal rather than formal models).

The weight on modeling the causes of effects in economics explains why many experimentalists who come from an economics tradition do not appear to be terribly interested in using their experiments to study a particular single cause-and-effect relationship in isolation but instead typically study a host of predicted relationships from some existing theory, as discussed in Chapter 6. These experimentalists usually begin with a formal model of some process, derive a number of predictions from that model, and then consider whether the behavior of subjects is in line with these predictions (or not) in their experiment. To researchers who have been trained to think of experiments as single tests of isolated cause-and-effect relationships as in the so-called classical experiment, these experiments

appear wrongheaded. But this failure is one of understanding, not of a method, which we hope our discussion of the formal theory approach to causality in this book will help reduce.

2.3.2 Effects of Causes and Inductive Research

However, not everyone agrees with Heckman's emphasis on theoretical models and the causes of effects. In his critique of Heckman's essay, Sobel (2005, p. 103) argues that many scientific questions are not causal, but purely descriptive. He remarks that "NASA . . . crashed a probe from the Deep Impact spacecraft into comet Tempel1 with the objective of learning more about the structure and composition of cometary nuclei." Sobel continues by pointing out that modeling the causes of effects is not important unless the effects of causes are sizable, noting that studying the causes of global warming is important because of the effects of global warming.

A lot of political science quantitative research – we would say the modal approach – is not so much into modeling or thinking beyond causality but instead focuses on investigating the effects of particular causes. Sometimes this activity is advocated as part of an effort to build toward a general model of the causes of effects, but usually if such a goal is in a researcher's mind, it is implicit. In experimental research, Gerber and Green (2002) advocate this approach in their call for use of field experiments to search for facts, as we discuss further later. Gerber and Green contend that experiments are a particularly useful way to discover such causal relationships, more useful than research with observational data. Experimentalists who have been largely trained from a statistical background and some political psychologists also tend to take this approach. The implicit idea is that eventually systematic reviews would address how these facts, that is, causes, fit together and help us understand the causes of effects.

Is there a "right" way to build a general model of the causes of effects? Morton (1999) maintains, as do we, that both approaches help us build general models of the causes of effects. Moreover, as Sobel holds, sometimes purely descriptive studies, which are not interested in causal questions, are useful. But it is a mistake to think that piecemeal studies of the effects of causes can be effectively accomplished without theorizing, just as it is a mistake to think that general models of the causes of effects can be built without piecemeal studies of effects of causes in the context of the models. To make this point, we explore how piecemeal studies of the effects of causes and approaches to building models of the causes of effects work in this and the following chapters.

2.3.3 An Example: Information and Voting

To illustrate how causal inference in political science research is conducted, we focus on a research area that has received significant attention, using both observational and experimental data and from researchers who use methods from a range of disciplines: What is the causal effect of information on voting behavior? What are the causes that determine how individuals vote in elections? We later elaborate on the nature of the research questions in terms of causality.

The Effects of a Cause Question

Elections often involve fairly complicated choices for voters. Even in simple two-candidate contests, voters vary in the degree over which they know the policy preferences of the candidates and how the candidates are likely to govern if elected. When elections involve more than two candidates or are referenda over specific legislation, voters' information about the consequences of their choices also varies. What is the effect of information about choices in elections on how voters choose? We know that uninformed voters are more likely to abstain; Connelly and Field (1944), in one of the first survey analyses of the determinants of turnout, found that nonvoters were two-thirds more likely to be uninformed about general political matters than those who participated. But, as Connelly and Field noted, the effect they discovered may simply reflect the fact that nonvoters are also less educated. Connelly and Field could not conclude that a lack of information caused voters to abstain. Much subsequent research has reported that this relationship is robust across election contests and years of study. Are these individuals not voting because they are less educated and, as a consequence, choosing to be uninformed because they are not voting or are they uninformed because they are less educated and, as a consequence, choosing not to vote? Or is there another factor, such as cognitive abilities or candidate strategies, that affects both whether someone is informed and whether they vote or not?

Furthermore, what is the effect of information on voting choices if voters do participate? Some uninformed individuals do vote. Do less informed voters choose differently than more informed voters who are similar in other ways, choosing different candidates as Bartels (1996) contends? Or do uninformed voters choose "as if" they are informed using simple cues like party labels or poll results, as argued by a number of scholars?[3] How much information do voters need to make "correct" decisions (decisions they would make if they were fully informed)? Can voters use simple cues

---

[3] See, for example, Berelson et al. (1954); McKelvey and Ordeshook (1985); Page and Shapiro (1992).

and cognitive heuristics as described by Kahneman et al. (1982) to make "correct" decisions? If uninformed voters would choose differently if they were fully informed, then does the distribution of information affect the ability of different voters to have their preferences affect electoral outcomes resulting in election outcomes that are biased in favor of the preferences of other, more informed voters? The answers to these questions are fundamental for understanding how electoral processes work and how elections translate voter preferences into outcomes. All of these answers hinge on how information influences voter choices, a question that turns out to be extremely difficult to determine and the subject of much continuing controversy.[4]

The Causes of an Effect: Questions and Theories of Voting

Furthermore, the relationship between information and voting is highly relevant to the task of building a general model of turnout and voting behavior in elections. Why do people vote? What determines how they vote? There are a number of competing explanations for voting behavior, most of which have specific implications for the relationship between information and voting. We explore the main ones because they are relevant for some of the examples that we use throughout this text.

The Expressive Voter. One explanation of how voters choose in elections is that voters choose whether to participate and how they vote for expressive purposes, which we label the Expressive Voter Theory.[5] Voters receive some value from participation and expressing their sincere preferences, and this induces them to do both. A version of this theory argues that one implication is that the more informed voters are, the more they are likely to participate in elections because they receive more utility from expressing preferences when they are informed about the choices.[6] Expressive voters are also predicted to make the choice that their information leads them to believe is ex ante their most preferred choice.

The Cognitive Miser. An explanation of how voters choose from political psychology is the view that voters are "limited information processors" or "cognitive misers" and make voting decisions based on heuristics and cues

---

[4] Contrast, for example, the conclusions of Bartels (1996), Lau and Redlawsk (2001), and Sekhon (2005) on whether uninformed voters vote "as if" they are informed and the literature reviewed on this subject. We address the reasons for these different conclusions subsequently.

[5] See Schuessler (2000), for example.

[6] See Matsusaka (1995).

as described earlier. These heuristics may lead to more informed choices with limited information or they may lead to systematic biases in how voters choose. As Lau and Redlawsk (2001, p. 952) remark: "Heuristics may even improve the decision-making capabilities of some voters in some situations but hinder the capabilities of others." Thus, the theory contends that how voters use these cognitive heuristics and whether they can lead to biased outcomes influences how information affects voters' choices. We label this the Cognitive Miser Theory.

The Primed, Framed, or Persuaded Voter. An extension of the Cognitive Miser Theory is the view that in politics, because voters are cognitive misers they can be easily influenced by information sources such as campaign advertising and the news media. That is, as Krosnick and Kinder argue, one heuristic that voters might use "is to rely upon information that is most *accessible* in memory, information that comes to mind spontaneously and effortlessly when a judgement must be made" (1990, p. 499, emphasis in the original). Because information comes to voters selectively, largely through the news media or advertising, biases in this information can have an effect on voter behavior. The contention is that the news media, by choosing which stories to cover and how to present the information, can "frame" the information voters receive, "prime" them to think about particular issues, or "persuade" voters to value particular positions, such that they are inclined to support political positions and candidates.

Chong and Druckman (2007) review the literature on framing and explain the distinctions among framing, priming, and persuasion as used in the psychology and communications literatures. Loosely, framing effects work when a communication causes an individual to alter the weight he or she places on a consideration in evaluating an issue or an event (e.g., more weight on free speech instead of public safety when evaluating a hate group rally), whereas priming in the communication literature refers to altering the weight attached to an issue in evaluations of politicians (e.g., more weight on economic issues than on foreign affairs in evaluating the president). Persuasion, in contrast, means changing an actual evaluation on a given dimension (e.g., the president has good economic policies). Thus, the theory argues that biases in the content of the information presented to voters and differences in presentations of the information can bias how voters choose in elections.

Effects of Negative Information. One particular aspect of information during election campaigns has been the subject of much disagreement in the

political behavior literature – the effects of negative campaign advertising. Ansolabehere and Iyengar (1997) suggest that some advertising can actually decrease participation. Specifically, they argue that negative advertising actually demobilizes voters by making them apathetic. The exposure to negative advertising, according to this theory, weakens voters' confidence in the responsiveness of electoral institutions and public officials generally. The negative advertising suggests not only that the candidate who is the subject of the negative ads is not someone to trust, but also that the political system in general is less trustworthy. Negative advertising then makes voters more negative about politics, more cynical, and less likely to participate. In contrast, others such as Lau (1982, 1985) have argued that negative advertising actually increases voter participation because the information provided can be more informative than positive advertising. The debate over the effects of negative advertising has been the subject of a large experimental literature in political science and is also a case for which a notable number of observational studies exist that use experimental reasoning. We discuss some examples from this literature.

The Pivotal Voter. An alternative theory of voting from political economics is what we label the Pivotal Voter Theory. In this model, voters' choices, whether to turn out and how to vote, are conditioned on being pivotal. That is, whether or how an individual votes does not matter unless his or her vote is pivotal. So when choosing whether and how to vote, an individual votes "as if" he or she is pivotal and does not vote at all if the expected benefits from voting (again conditioned on pivotality) are less than the cost. In a seminal set of papers, Feddersen and Pesendorfer (1996) apply the pivotal voter model to understand how information affects voters' choices. They show that the theory predicts that uninformed voters may be less likely to vote than informed voters if they believe that informed voters have similar preferences because they wish to avoid affecting the election outcome in the wrong direction. Moreover, the less informed voters may vote to offset partisan voters whose votes are independent of information levels. According to the theory, then, it is possible that less informed voters may purposely vote against their ex ante most preferred choices to offset the partisan voters. These particular predictions about how less informed voters choose has been called by Feddersen and Pesendorfer the Swing Voter's Curse.

The Voter as a Client. Electoral politics in many developing countries has been theorized by comparative politics scholars as a clientelist system. Clientelism is when the relationship between government officials and voters is

characterized as between a rich patron who provides poor clients with jobs, protection, and other specific benefits in return for votes. Thus, in such systems, campaign messages are about the redistributive transfers that the elected officials plan to provide to their supporters. Voters choose candidates in elections that they believe are most likely to provide them with the most transfers. Information about what candidates will do once in office in terms of such transfers can thus affect voters' choices to the extent that they value the transfers.

Of course, because voting is a fundamental part of political behavior and has been the subject of extensive theoretical examination, other theories exist of how people vote, such as group models of voting described by Feddersen and Sandroni (2006), Morton (1987, 1991), Schram (1989), and Uhlaner (1989). We focus on the aforementioned theories because they have been addressed using experimental work that we use as examples in this chapter.[7]

The Broader Implications

Evaluating the causal effect of information on turnout and how individuals vote in the ballot booth provides evidence on whether these particular implications of the more general models of the causes of voting are supported. Such research, combined with evaluations of other implications of these theories, works to determine what causes turnout and what causes how voters choose in the ballot booth.

Furthermore, the answers to the questions of effects of a cause and the causes of an effect also affect how we answer other important policy questions about elections and campaigns. For example, how do campaign advertisements influence voters' choices (if at all)? Do ads need to be substantively informative to influence uninformed voters to choose as if they are informed or can voters use simple ads that mention things like party or other simple messages to make "correct choices?" Is it important that the media provide detailed substantive information on candidate positions? Can biased media reporting on candidate policy positions influence voters? How important are debates in which candidates discuss substantive issues in the electoral process? These policy questions depend not only on the particular causal effect of information on voting but also how we answer the questions about why voters turn out and the determinants of how they vote.

These questions are also useful for an exploration of how causality is investigated in political science using both experiments and nonexperimental

---

[7] Feddersen et al. (2009) provide an interesting experimental test of a theory of voting related to the Feddersen and Sandroni model of ethical voting.

empirical studies since many researchers have tackled them using both types of data, including even natural experiments. Thus, we can use these studies as examples in our exploration. However, it is important to recognize that the examples are not necessarily ideal cases; that is, researchers have made choices that may or may not have been optimal given the question at hand, as we note. The examples are meant as illustrations of how actual research has been conducted, not always as exemplars for future research.

2.4 Setting Up an Experiment to Test the Effects of a Cause

2.4.1 The Data We Use

The Data Generating Process

We begin our study of causality with the effects of a cause question. We use our example of information and voting as an illustration. We also show how experimental reasoning works within our example. We consider an election in which there are two or more options before voters who must choose only one. The election might be for President of Mexico, Mayor of Chicago, a member of the British Parliament, a referendum on a policy proposal, or an election created by an experimentalist in a laboratory (where what we mean by a laboratory experiment is defined more precisely later). That is, it may be the case that individuals, which we call subjects, have been brought to the laboratory and asked to choose between a set of candidates in an election set up by a researcher. The candidates could also be subjects or they might be artificial or hypothetical actors. Voters face a choice over whether to vote, and if they vote, which candidate to vote for. We think of the data generated by the election as created by a general data generating process (DGP) that provides the source for the population of data that we draw from in our research.

Definition 2.1 (Data Generating Process or DGP): *The source for the population of data that we draw from in our empirical research.*

The Target Population

The DGP is the source for lots of populations of data, not just one election. When we think of the DGP we think of data generated in all the countries of the world (and possibly outside our world). But we are typically interested in just a subset of the data that is generated. What population are we interested in? We have to choose a particular target population to study. If the election we are studying is a U.S. presidential election, then our target population includes the data generated by that election. Alternatively, if we are conducting an election in a laboratory, then the target population would

be the population of observations that are generated by such an election set up by the researcher in the laboratory. When we choose to study a particular election or set of elections, we effectively choose a target population. In our analyses, we typically use a sample of data drawn from the target population of data, which is a subset of the target population. The extent that the sample represents the target population is a question of statistical validity and is addressed in Chapter 7.

Definition 2.2 (Target Population): *The population of observations generated by the DGP that an empirical researcher is addressing in his or her analysis.*

2.4.2 What Is an Experiment?

Intervention and Manipulation in the DGP
In an experiment, the researcher intervenes in the DGP by purposely manipulating elements of the environment. A researcher engages in manipulations when he or she varies parts of the DGP so that these parts are no longer naturally occurring (i.e., they are set by the experimenter). We might imagine an experimenter manipulating two chemicals to create a new one that would not naturally occur to investigate what the new chemical might be like. In a laboratory election experiment with two candidates, a researcher might manipulate the information voters have about the candidates to determine how these factors affect their voting decisions. In both cases, instead of nature choosing these values, the experimenter chooses them. Our laboratory election is a particular type of experiment in the social sciences in which subjects are recruited to a common physical location called a laboratory and the subjects engage in behavior under a researcher's direction at that location.

Definition 2.3 (Experiment): *When a researcher intervenes in the DGP by purposely manipulating elements of the DGP.*

Definition 2.4 (Manipulation in Experiments): *When a researcher varies elements of the DGP. For a formal definition of the related concept, manipulated variable, see Definition 3.3.*

Definition 2.5 (Laboratory Experiment): *Where subjects are recruited to a common physical location called a laboratory and the subjects engage in behavior under a researcher's direction at that location.*
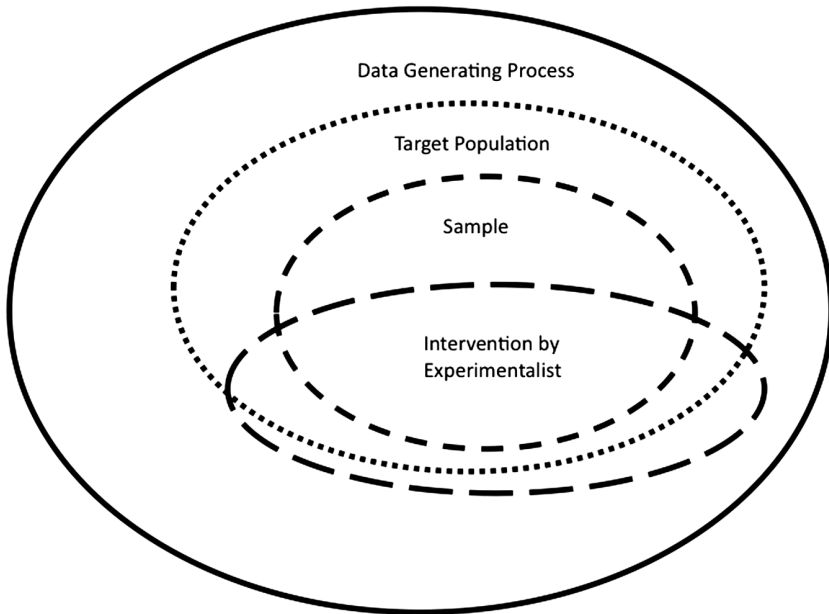
Figure 2.1 Relationships between the Data Generating Process, Target Population, Sample for Study, and Experimental Intervention.

The intervention and manipulation of the experimenter ideally principally affect the target population in the study (and the sample drawn from that population that is studied by the researcher). However, the intervention and manipulation may also affect other parts of the DGP by affecting choices of individuals outside of the target population. For example, when a researcher pays subjects for their participation in an experiment, the payments may affect the income and choices of individuals who are not part of the target population of the experiment as the subjects spend the money given to them. Figure 2.1 above illustrates the case in which the intervention affects observations outside the target population (and outside the sample drawn by the experimentalist).

Experimental Control

Confounding Factors. Experimenters worry (or should worry) about factors that might interfere with their manipulations. For example, trace amounts of other chemicals, dust, or bacteria might interfere with a chemist's experiment. That is, the chemist may plan on adding together two chemicals, but when a trace amount of a third chemical is present, his or her manipulation is not what he or she thinks it is. Similarly, if a researcher is manipulating the

information that voters have in a laboratory election, factors such as how the individual receives the information, the individual's educational level, how much prior information the individual has, the individual's cognitive abilities, the individual's interest in the information, or the individual's mood at the time he or she receives the information all may interfere with the experimenter's ability to manipulate a voter's information. The researcher intends to manipulate voter information but may or may not affect voter information as desired if these confounding factors interfere.

Definition 2.6 (Confounding Factors): *Factors that can interfere with the ability of an experimentalist to manipulate desired elements of the DGP.*

Early experimenters were aware of these possible confounding factors. As a result, they began to control possible confounding factors when they could. Formally, a researcher engages in control when he or she fixes or holds elements of the DGP constant as he or she conducts the experiment. A chemist uses control to eliminate things that might interfere with his or her manipulation of chemicals. In a laboratory election, if a researcher is manipulating the information voters have about candidates, the researcher may want to hold constant how voters receive the information and how much other information voters have so that the researcher can focus on the effects of information on how voters choose in the election.

Definition 2.7 (Control in Experiments): *When a researcher fixes or holds constant elements of the DGP to better measure the effects of manipulations of the DGP.*

Observable Versus Unobservable Confounding Factors and the Advantage of the Laboratory. The confounding factors can be of two types: observable and unobservable. Observable factors are simply things that the researcher is able to measure with only random error. For example, in a laboratory election, how the individual receives the information or the individual's educational level are things the researcher can measure arguably with only random error. In contrast, the individual's interest in the information or mood may be something that the researcher cannot observe with confidence. We would call such a factor an unobservable factor. What is observable and unobservable depends on the circumstances of the manipulation and the target population studied. That is, some potential confounding factors such as an individual's educational level may be observable in an experiment conducted with voters participating in a U.S. presidential election as well as in a laboratory election, but it might be easier to observe how much prior

information voters have in a laboratory election than in an experiment that is part of a U.S. presidential election. Thus, in the first case, prior information may be observable, but in the latter case, it is unobservable.

As a consequence, to facilitate control, most early experiments in the social sciences, as in the physical sciences, were conducted in laboratories. In the laboratory, many confounding factors can be made observable and the experimentalist can then control for their possible interference. As noted in the preceding example, in a laboratory election a researcher can, by creating the election that takes place, make observable voters' prior information, allowing the researcher to better control voters' prior information, which may be unobservable outside of the laboratory.

Definition 2.8 (Observable Confounding Factors): *Confounding factors that a researcher is able to measure in the target population with only random error given the experimental manipulation.*

Definition 2.9 (Unobservable Confounding Factors): *Confounding factors that a researcher cannot measure with any confidence in the target population given the experimental manipulation.*

Baselines in Experiments. One method of controlling confounding variables is to compare experimental results to outcomes in which manipulations do not occur but all other observable conditions are identical. That is, if all the other conditions are held constant – are identical – and the only difference between the two outcomes (the outcome when the manipulation did not occur and the outcome when the manipulation did occur) is the experimental manipulation, then the researcher can argue that the effect he or she is measuring is truly causal; that is, the manipulation has caused any differences between the two outcomes. Oftentimes experimentalists call the outcome in which a manipulation did not occur the "control" and the experiment a "controlled experiment." However, because control is a more than just a comparison, but involves other ways that experimentalists attempt to control confounding variables, we label such a comparison a "baseline comparison." Also, the word control is used in observational studies in the same general sense: as a method of holding constant the effects of possible confounding variables. We discuss baselines more expansively in Section 8.3.

Definition 2.10 (Baseline): *A manipulation in an experiment designated by a researcher as being particularly relevant for comparisons. For a formal definition of the related concept, baseline treatment, see Definition 8.7.*

Random Assignment

Moving Out of the Laboratory. As Shadish et al. (2002; hereafter SCC) observe, when experimentation moved out of the laboratory and expanded to disciplines such as agriculture, public health, education, and so forth, researchers were no longer able to control adequately aspects of the DGP that might interfere with their manipulations. Researchers could not always find observations with identical observables, and they encountered more unobservable possible confounding factors. A field experiment using human subjects – the name probably comes from agricultural use – is a researcher's intervention that takes place in subjects' natural environments and the researcher has only limited control beyond the intervention conducted.

Definition 2.11 (Field Experiment): *Where a researcher's intervention takes place in subjects' natural environments and the researcher has only limited control beyond the intervention conducted. Usually the relationship between the researcher and the subject is conducted through variables outside of the researcher's control.*

In field experiments in agriculture, it was difficult to use control to account for differences in soil, farming abilities, and so on. It was not possible to find two fields with exactly the same observable conditions and unlikely that unobservable conditions were the same, and researchers expected that these factors could confound the manipulations. Thus, when comparisons were made between the baseline outcome that resulted from no manipulation and outcome that occurred as a consequence of a manipulation, the experimenter could not be sure if the difference was due to the manipulation or to the differences in the fields that he or she could not control.

In field experiments in public health and education, researchers similarly lost the ability to control for variables that could confound their manipulations. It was not possible for them to compare individuals who were exactly the same, living in exactly the same environment, eating the same food, with the same prior health conditions, psychological makeup, or cognitive abilities. Similarly, if a researcher wished to manipulate the information voters have in an election that is naturally occurring as part of the DGP, then the researcher no longer has as much control over how the voters receive information and how much other information voters have in the same way as the researcher can control the information in the laboratory. That is, suppose the information is provided through a mailing about candidate positions on issues. Some voters may not receive the mailing because of mistakes in addresses, others may not check their mail, and still others may

throw the mailing away without reading it. Furthermore, some voters may already know the information. These disconnects would occur to a much smaller extent in the laboratory.

As a result of the inability to control factors outside the laboratory and the difficulty in comparing human subjects, researchers in agriculture, biomedicine, and social sciences began to develop techniques such as random assignment as substitutes. Random assignment is when the researcher uses a randomization mechanism to assign subjects to manipulations, one of which might be a baseline manipulation. In our simple example, the researcher may randomly assign some subjects to receive the information manipulation about the candidates and others to receive no information (a baseline).

Definition 2.12 (Random Assignment): *When a researcher uses a randomized mechanism to assign subjects to particular manipulations in the experiment to better measure the effects of manipulations of the DGP.*

Is Random Assignment Essential for an Experiment? As shown in Section 5.2.2, random assignment can facilitate the ability of researchers to establish causal inferences. Essentially, because the information is randomly assigned across subjects, then the factors that might interfere with the effects of the manipulation, such as whether the subjects actually received the information or already knew the information, are in expectation mitigated (the effects do not disappear, but on average are controlled assuming the randomization is effective). The importance of random assignment for experiments conducted outside of the laboratory in public health, education, and similar disciplines led some to restrict the definition of an experiment using human subjects to one in which random assignment is used. For example, SCC define an experiment explicitly as an intervention that uses random assignment, and an intervention that does not is defined as a quasi-experiment because their focus is largely on experiments conducted in these disciplines. Many political scientists have adopted the same convention. Certainly SCC are correct to say, as we explore later in this book, that when one compares two experiments conducted outside the laboratory that are exactly alike except that manipulations in one experiment are assigned randomly and in the other they are not, the one in which the manipulations are assigned randomly is likely to do better in establishing causal inferences than the other, and can certainly do no worse.

However, if we were to compare a laboratory experiment that did not use random assignment but the researcher engaged in significant control over the elements of the DGP to an experiment outside the laboratory in which

little control is exercised but random assignment is used, the answer is not so clear. Any experiment with random assignment does not always make "better" causal inferences than any experiment without random assignment. Why? There are two reasons. First, control also facilitates causal inferences, as we discuss in Section 4.1.1. For example, in some laboratory experiments, researchers use what is called a within-subjects design (defined and discussed in Section 3.3.3), which can have advantages over simple random assignment in establishing causality because the same subjects experience all manipulations even if everything else about an experiment is held constant. Subjects serve as their own baselines. Random assignment implies that subjects in expectation have the same probability of experiencing a manipulation, but a within-subject design makes that probability equal to 1 for both manipulations across all subjects.

Second, operationalizing random assignment in experiments is not simple and involves a number of decisions, about what to randomize, across what groups of subjects, and so on, that can affect the value of the inferences made through random assignment. Furthermore, when researchers conduct experiments, especially when the experiments are conducted in the field, issues of response and compliance become important. Nonresponse is when a subject's choices, given manipulations, are not observable, and noncompliance occurs when a subject fails to comply with the manipulation given by the researcher. Random assignment, particularly in field experiments, is thus rarely as ideal for establishing causal inferences as the statistical theory that underlies it would suggest. Thus, both control and random assignments are methods used to deal with factors that can interfere with manipulations; neither is perfect, but both are extremely powerful.

Definition 2.13 (**Nonresponse**): *Nonresponse is when a subject's choices, given manipulations, are not observable.*

Definition 2.14 (**Noncompliance**): *Noncompliance occurs when a subject fails to comply with the manipulation given by the researcher.*

Consider some well-known deliberative polling experiments (see Fishkin, 1991, 1993 to 1997; Luskin et al. 2002). In these experiments, a random sample of subjects was recruited to participate in an event to discuss and deliberate public policy on a particular issue. The early experiments suffered from the lack of an explicit baseline sample, noncompliance when subjects who were selected to attend did not, and nonresponse when subjects who attended did not respond to surveys after the event. As a result, many have

argued that these events are not experiments, labeling them quasi-experiments, as in the discussion by Karpowitz and Mendelberg (forthcoming). We agree that the methodological concerns of the critics are justified. The design of the deliberative polls makes it difficult to draw causal inferences about the effects of deliberation on public opinion. However, not all of these experiments lacked a baseline group and an attempt at random assignment. For example, Barabas (2004) reports on a deliberative poll event in which a baseline group was surveyed and random samples of subjects were recruited to both a baseline group and a group that participated in the deliberative poll. However, the random assignment was problematic because some subjects (fewer than 10%) were recruited to participate independently by interest groups, some subjects chose not to participate (noncompliance), and others did not respond when surveyed post-poll (nonresponse). Barabas labels the experiment a quasi-experiment as a consequence of these problems with the attempt at random assignment despite the efforts of the researchers to draw random samples for both the baseline and manipulated groups. Since almost all field experiments suffer from similar problems in implementing random assignment, it would seem that a strict interpretation of what is an experiment along these lines would ultimately mean that only a few real field experiments exist in political science.

Some important and useful experiments have been conducted that do not use random assignment or baselines or that fail to fully implement random assignment, yet they have added significantly to our understanding of political behavior and institutions just as many experiments in which the researcher has little control over variables not manipulated also have provided useful knowledge. The fact that a study does not include randomization or baselines or the randomization suffers from problems, in our view, does not make it less of an experiment, just as an experiment in which control is minimal is not less than an experiment. As we explain in Section 8.2.4, we think it is important not to confound definitions of experiments with normative views of desirable properties because what is desirable in an experiment depends on the research goal – what the researcher seeks to learn – as well as the opportunities before the researcher. What is ideal in an experiment also depends on where it is conducted. In field experiments, random assignment can be extremely valuable, although difficult, because control is less available; in the laboratory, the opposite relationship holds although both control and random assignment can be much easier to implement. It would be unreasonable for us to define interventions outside the laboratory, where there are disconnects between manipulations, and what happens to subjects because of a lack of control or problems with the

implementation of random assignment as not really experiments, just as we think it is unreasonable to define interventions without random assignment and baselines as not really experiments. Thus, we define experiments broadly following the traditional definition: an experiment is simply an intervention by a researcher into the DGP through manipulation of elements of the DGP.[8] We further define control and random assignment with or without baselines as usual and important tools by which a researcher can more fruitfully make causal inferences based on his or her interventions. But we recognize that both control and random assignment are rarely implemented perfectly, especially when the experiment is conducted in the field, and thus defining an experiment by whether it contains either one is not useful.

2.4.3 Examples of Information and Voting Experiments

In the Appendix to this chapter contains seven examples of experiments on the relationship between information and political choices. In some cases, the subjects voted in an election conducted by the experimenters; in other cases, subjects reported on their preferences over choices that were presented to them as candidates or choices, sometimes in a hypothetical election, sometimes in an upcoming naturally occurring election in which the subjects would vote or had voted. In some cases turnout decisions, rather than choices between candidates or parties, were measured or surveyed. In all the examples, the experimenters manipulated or attempted to manipulate the information the subjects possessed about the choices before them, how the information was presented to the subjects, or both. All seven used some form of random assignment to manipulations and comparison of manipulations.

The examples, however, illustrate the wide variety of experimental approaches used in political science. Three of the example experiments were conducted during elections in the field. Example 2.1 presents an experiment by Gerber et al. (2007) in which they provided subjects with free newspaper subscriptions during a Virginia gubernatorial election; Example 2.2 concerns an experiment by Wantchekon (2003) during a presidential election in Benin in which he manipulated the campaign messages used by some of

---

[8] Our definition of an experiment is also traditional in that researchers used experimentation for many years before the advent of random assignment as a tool in establishing causal inferences in the early twentieth century. If we label interventions into the DGP as nonexperiments if they do not use random assignment, many famous and infamous research trials would be considered nonexperiments, such as Edward Jenner's research leading to the smallpox vaccine and the Tuskegee syphilis study discussed in Section 11.4.1.

the political parties; and Example 2.3 discusses an experiment by Clinton and Lapinski (2004) during the 2000 U.S. presidential election, in which Clinton and Lapinski showed subjects negative campaign advertisements. Clinton and Lapinski's experiment is an Internet survey experiment because it is embedded in a survey and conducted via the Internet. We discuss these particular types of experiments more expansively in Section 8.2.1.

The other four examples are laboratory experiments. However, they vary across important dimensions. Two were conducted by political psychologists involving hypothetical candidates (Example 2.4, experiments conducted by Kulisheck and Mondak (1996), Canache et al. (2000), and Mondak and Huckfeldt (2006), and Example 2.5, conducted by Mutz (2007)). Mondak and his coauthors varied how much information subjects had about candidate qualities, whereas Mutz varied the visual mechanism by which subjects learned about candidates. The remaining two examples were conducted by political economists in which subjects chose in a laboratory election and were given payments based on which choices won (Example 2.6, conducted by Battaglini, Morton, and Palfrey, and Example 2.7, conducted by Dasgupta and Williams [2002]). In Dasgupta and Williams's experiment, the choices before voters were called candidates, and voters were told they were voting in an election, whereas in Battaglini, Morton, and Palfrey's (2008, 2010) study, subjects were asked to guess the color of an unseen jar given information provided to them and the "winner" was the color that received the majority of guesses. Some of the experiments were conducted via computer networks. (Battaglini et al., Dasgupta and Williams, and Mondak et al. used computers in a laboratory; Clinton and Lapinski's experiment was conducted via the Internet). Mondak et al. and Mutz also used other methods to measure how subjects used or responded to information: Mondak et al. measured the time taken by subjects to respond to various questions and Mutz measured skin reactions to visual presentations of information.

We can also see tremendous variation in the control used in the experiments. In Examples 2.1, 2.3, and 2.5, Gerber et al., Clinton and Lapinski, and Mutz, respectively, designate one of the manipulations as a baseline manipulation. In other examples, although comparisons are made, the manipulation that one would call a baseline is not so obvious, as in Dasgupta and Williams's and Wantchekon's experiments. In the field experiments, the researchers generally had little control over many possible confounding variables. For instance, in Example 2.1, Gerber, Kaplan, and Bergan have little control over what is reported in the newspapers about the election, whether subjects read the newspaper articles about the election, and to some extent whether their access to the newspaper is manipulated. In contrast, in the laboratory experiments in Example 2.5, the subjects came to Mutz's

laboratory and watched videos she had prepared in a controlled setting. In studies by Dasgupta and Williams and Battaglini et al., the researchers used financial incentives in an attempt to control subjects' preferences over their choices in the elections, which was not possible in the other five examples. That is, in the other five examples, the researchers must account for partisan preferences held by the subjects that may also affect their choices but cannot control them explicitly.

The seven examples also differ in the types of subjects used. The field experiments used residents in the localities where they were conducted, whereas the laboratory experiments generally used students, with the exception of Mutz's experiment which used nonstudents recruited via employment agencies or civic groups. Clinton and Lapinski used a national sample from an Internet survey organization. Some of the laboratory experiments that used students drew subjects from multiple universities: Mondak et al. used students from the United States, Mexico, and Venezuela, and Battaglini et al. used students from Princeton University and New York University. We discuss the advantages and disadvantages of different types of subjects in Chapter 9.

Finally, all seven examples reference to varying degrees one or more of the theories of voting mentioned in Section 2.3.3. Gerber et al., Mondak et al., and Mutz reference either explicitly or implicitly Cognitive Miser views of voting in which information primes, frames, or persuades voters. Clinton and Lapinski contend that their experiment provides important new evidence on negative advertisements, while Wantchekon argues that his work similarly informs our understanding of how clientelism works. Dasgupta and Williams and Battaglini et al. relate their experimental work to pivotal voter models. And both Gerber et al. and Battaglini et al. discuss expressive theories of voting. We address these and other variations in the examples (and additional examples presented later) throughout the text.

2.4.4 What Is Not an Experiment?

Qualitative Research and Traditional Surveys

Although our definition of experiments is encompassing, it excludes other research with human subjects in the social sciences such as interviews and qualitative, soak-and-poke, political science research that aims not to intervene in the DGP but to measure and observe how the DGP operates through close interaction with human subjects. Manipulations that occur in these types of research studies are not generally purposeful, but accidental. Whether, of course, it is possible to observe in close interaction without manipulating the DGP is an important issue in such research, but the overall

goal, as we understand it, is access to human subjects in the DGP as if the researcher were not there, rather than to manipulate the DGP. If qualitative researchers see themselves as intervening with the purpose of altering the DGP, we call that research an experiment. Similarly, a traditional survey is not an experiment because the goal of the researcher is to measure the opinion of the respondents, not to intervene or manipulate elements of the DGP that affect these opinions. When a researcher does purposely attempt to use a survey to manipulate elements of the DGP that theoretically affect respondents' opinions, we call this an experiment. We discuss experiments in surveys more expansively in Section 8.2.1.

Note that we recognize that the goal of many experimental manipulations is to better measure political behavior or preferences, as in many of the political psychology experiments which use implicit messages in an attempt to better measure racial prejudices (see, for example, Lodge and Taber, 2005). Yet the means of achieving the goal is through manipulation, not passive observation, which makes this research experimental rather than observational, in our view.

Natural and Policy Experiments and Downstream Benefits of Experiments

Sometimes nature acts in a way that is close to how a researcher, given the choice, would have intervened. For example, hurricane Katrina displaced thousands of New Orleans residents and changed the political makeup of the city, as well as having an impact on locations that received large numbers of refugees. Although no political scientists we know would wish such a disaster to occur in a major city, many find the idea of investigating the consequences of such a manipulation an exciting opportunity to evaluate theories of how representatives respond to changing constituencies, for example. Katrina was an act of nature that was close to what a political scientist would have liked to have done if he or she could – intervening and changing the political makeup of several large U.S. cities such as New Orleans, Houston, and Atlanta.

Natural manipulations might also occur in our information and voting example. For instance, in the case where the mailing described earlier in a naturally occurring election is provided without input by a researcher, then it is a natural manipulation. When natural manipulations occur, sometimes researchers argue that the manipulation is "as if" an experimentalist manipulated the variable. The researcher often calls the manipulation a "natural experiment," although the name is an oxymoron because by definition an experiment cannot be a situation where the DGP acts alone and thus we do not call these experiments, according to our definition. The researcher is contending that nature has two sides: the side that generates most data, and

then the interventionist side that occasionally runs experiments like aca-
demics, messing up its own data generating process. Even though in this case
the researcher is not doing the intervening, the approach taken with the data
is as if the researcher has. When does it make sense for a researcher to make
such a claim and approach his or her observational data in this fashion? The
answer to this question is complicated and we address it fully in Section 5.4.3.
Example 2.8 in the Appendix presents a study of a natural experiment
involving the effect of information on voter turnout by Lassen (2005).

Definition 2.15 (Natural Experiment): *Nonexperimental or observational
data generated by acts of nature that are close to the types of interventions or
manipulations that an experimentalist would choose if he or she could.*

A special type of natural experiment occurs when government offi-
cials manipulate policies. For example, in Lassen's experiment, the natural
manipulation occurred when government officials varied the ways in which
public services were provided. Similarly, De La O (2008) exploits govern-
mental policy changes in Mexico to consider how different governmental
services impact voter participation. We call such a manipulation a policy
experiment when it is undertaken by government officials without academic
involvement.

Definition 2.16 (Policy Experiment): *A field experiment in which a gov-
ernment agency or other institution chooses to intervene and act "like an
experimentalist."*

More recently, governments and other nonacademic institutions have
formed collaborations with academic researchers to conduct experimental
manipulations. An example of such a collaboration is provided in Exam-
ple 12.1, where Olken (2008) collaborated with officials to manipulate the
mechanisms by which villages in Indonesia made decisions about which
public projects to fund. When such collaboration occurs and the researcher
is directly involved in consciously choosing the design of the manipulation,
the manipulation is an experiment, as we have defined. If the collabora-
tion does not involve a researcher in the design process but simply allows a
researcher the opportunity to gather data on a manipulation already planned
and designed for a different purpose, the research is a natural experiment
and not an experiment, as we have defined.

Occasionally researchers might use the manipulation of an experi-
ment conducted in the past to investigate either a new hypothesis or the

long-term implications of the original manipulation. Gerber and Green (2002) label such research downstream research and the results of these investigations the downstream benefits of experimentation. The use of previous experiments in this fashion certainly has advantages in the same ways that natural and policy experiments can be useful in empirical analysis. See Sondheimer (forthcoming) for a discussion of how researchers can benefit from prior manipulations.

Definition 2.17 (Downstream Benefits): *The benefits of analysis of prior experiments, either conducted by academics for research or conducted by government as a policy experiment.*

Computational Models and Simulations

Occasionally political scientists who use computational or agent-based models to numerically solve formal models of politics call the output "experiments," or others who run counterfactual analyses using parameters estimated from an empirical model call their analyses "experiments" (see, e.g., Kollman and Page, 1992). Computer simulations to solve formal models are aids in solving a model, not in "testing" the model with empirical data. These simulations are an extension of the researcher's brain. Similarly, counterfactual simulations using parameters from estimated empirical models are aids to understanding the empirical model estimated using either observational or experimental data and are an extension of the analysis of those data, not the generation of new experimental data. In experiments the subjects make "real" decisions and choices and are independent of the researcher, and new data are generated. The subjects are not simulating their behavior, but engaging in behavior. The environment created by the experimentalist is not an observational environment but it is real in the sense that real individuals are involved. Thus, simulations and experiments serve entirely distinctive purposes.[9]

Counterfactual Thought Experiments

Related to the use of computational or agent-based models to solve formal models are what have been called counterfactual thought experiments, in which nonformal theorists hypothesize the effects of situations in which

---

[9] Some experimentalists believe that this also means that it is important that the experimentalist not "hard-wire" the experiments by telling subjects how to choose or behave. However, there is no hard-and-fast rule on such suggestions, because in some cases doing so may be an important part of the experiment and is the subject of the experimental investigation.

one or more observables takes on values contrary to those observed (see, e.g., Tetlock and Belkin [1996] and Tetlock et al. [2006]). For example, what would have happened if Great Britain had confronted Hitler more before World War II? Again, these are not experiments as we have defined them because they are extensions of the researcher's brain – theoretical speculation about what would have happened historically if a variable had been manipulated.

2.4.5 Experimental Versus Observational Data

In most experimental research, the variation in the data is partly a consequence of the researcher's decisions before the data are drawn and measured. If we think of the DGP before or without intervention as nature acting alone, then the DGP after intervention is nature and the researcher interacting. We call the data generated by such intervention "experimental data." So for instance, in Clinton and Lapinski's experiment, Example 2.3, the 2000 presidential election without the experiment would be nature acting alone, but with the intervention of the researchers is nature and Clinton and Lapinski interacting. Data on the presidential election that do not involve such interaction are observational data (such as who the candidates were), but data generated through the interaction (such as how the subjects voted after having the campaign advertisements they saw manipulated by the researchers) are experimental data.

Definition 2.18 (Experimental Data): *Data generated by nature and the intervention of an experimentalist.*

Nonexperimental empirical research involves using only data drawn from the population in which all variation is a consequence of factors outside of the control of the researcher; the researcher only observes the subset of data he or she draws from the DGP but does not intervene in that process or if he or she does so, it is an accidental intervention, not purposeful. There are many observational studies of the 2000 U.S. presidential elections of this sort. This approach assumes, of course, that the researcher can and will choose to measure the data perfectly; clearly, choices made in measurement can result in a type of post-DGP intervention, but the data are still not experimental data because the data are generated without intervention.

Some distinguish between experimental and "naturally occurring" data. Others talk of the "real world" versus the experimental world. Such terms are misleading because nature is also involved in determining variation in

experimental data. Even in laboratory experiments, although the experimenter may intervene in the data generating process, the subjects in the experiment who make decisions are "real" and their decisions occur "naturally," albeit influenced by the experimental environment. Since as political scientists we are interested in human behavior, we should recognize that the humans participating in an experiment are as "real" as the humans not participating in an experiment.[10] A more neutral description of research using only data where the variation is a consequence of factors outside of the control or intervention of the researcher is research that uses observational or nonexperimental data; we use that terminology.

Definition 2.19 (**Nonexperimental or Observational Data**): *Data generated by nature without intervention from an experimentalist.*

2.5 Chapter Summary

Fundamentally, scientific research is about building and evaluating theories about the causes of effects. Political scientists are interested in studying why and how people vote, for example. One of the ways we build toward such theories and evaluate them is to study the effects of causes. Many theories of why and how people vote make causal predictions about how information, a cause (either in content or presentation), affects voters' choices, an effect. In this chapter we have reviewed some of these theories of voting and their predictions about the relationship between information and voting as an illustration. To evaluate theoretical predictions, or sometimes just hunches, about the relationship between information and voting, many political scientists have used experiments. We have discussed examples of these in this chapter and the examples are presented more fully in the Appendix.

In this chapter we have also surveyed the features of the experimental method used by researchers to address predictions. In summary, the standard use of the experimental method in political science typically involves the following four principal features:

1. Designating a target population for the experimental study,
2. Intervention and manipulation in the DGP,

---

[10] We discuss in Chapter 7 whether the experimental environment leads subjects to make choices they would not make if the same changes in their environment would occur via nature or the DGP, rather than through experimental intervention. Even if "being in an experiment" has such an effect, this does not mean that the subjects' choices are less real or less human. It means that we must understand that sometimes the aspects of the experiment itself that cause such effects are treatments that must be considered explicitly.

3. Controlling observable confounding factors, and
4. Random assignment of subjects to manipulations to control for unobservable confounding factors and observable confounding factors that are difficult to control for otherwise.

In Chapter 9 we comprehensively investigate the choice of a target population. But we have argued that the key component of experimentation is the second feature, intervention and manipulation in the DGP. Some research with human subjects, where such intervention and manipulation is not the intended goal, we do not include as experiments. Similarly, other research activities that have the label experiment, such as natural experiments and computer simulations, we also do not include because they do not involve intervention or manipulation in the DGP by a researcher.

Furthermore, we have argued that although the second two aspects – control and random assignment – are valuable features of experimentation, experiments can vary significantly in how these two components are manifested. Sometimes control is manifested in the use of baseline manipulations and a high degree of control over the environment of the experiment; other times control is only minimally used by an experimentalist. Correspondingly, sometimes random assignment is implemented nearly perfectly, in which all subjects comply with assignments and their choices are fully observable; other times experimentalists must deal with noncompliance with manipulations and an inability to observe subjects' behavior (nonresponse).

Because these two aspects of experimentation are not binary but are closer to continuous variables, we do not define experiments by whether they have a given level of control or random assignment. Instead, we argue that the degree that control and random assignment are used depends on the goal of the experiment, something that we explore more expansively in the chapters to come, with an extensive discussion of control in Chapter 4 and random assignment in Chapter 5. Before we turn to our detailed examination of these aspects of experimentation, we present the Rubin Causal Model, which is one of the main approaches that underlies causal inference in experimental political science. We discuss the second approach to causality in experimental political science, the Formal Theory Approach, in Chapter 6.

2.6 Appendix: Examples

Example 2.1 (Newspaper Field Experiment): *Gerber et al. (2007) report on an experiment conducted during the 2005 Virginia gubernatorial*

election, designed to see if biased information sources affected voter behavior.

Target Population and Sample: *Gerber, Kaplan, and Bergan selected a set of residents in Prince William County, Virginia, which is 25 miles from Washington, D.C. The subjects were selected about one month before the gubernatorial election in two waves from two different lists – a list of registered voters and a consumer database list. The registered voter list provided names of 54% of the first wave and 46% of the second wave; the consumer database provided the other names. Once these names were selected, the residents were surveyed in September 2005 and were asked if anyone in the household received either the* Washington Post *or the* Washington Times *newspapers. Respondents who answered "yes" or refused to answer any one of the questions in the survey were excluded from the study. The other questions were about the respondent's newspaper readership in general, the respondent's gender, whether he or she had voted in previous elections, and who she or he supported in the coming gubernatorial election.*[11] *This yielded a total of 3,347 subjects.*

Environment: *Gerber, Kaplan, and Bergan make use of the fact that Washington, D.C., has two national newspapers, the* Washington Post *and the* Washington Times; *the first is generally viewed as a liberal newspaper and the second is widely viewed as a conservative paper. Furthermore, researchers have found empirical evidence in support of the popular perceptions. Groseclose and Milyo (2005) compare the similarity of the experts used by media outlets with those cited by conservative and liberal members of Congress. From this comparison, they construct a measure of the ideological bias of newspapers. They find that the* Times *is the most conservative of the six papers they evaluate and the* Post *is much more liberal. Furthermore, the* Post *had endorsed the Democratic candidate for the Virginia gubernatorial election and the* Times *had endorsed the Republican. Thus, they have the opportunity to compare the effects of exposure on voting behavior in the Virginia gubernatorial election of two apparently very different news sources where the framing of stories, priming on issues, and persuasive efforts are arguably biased toward different candidates.*

Procedures: *The subjects were randomly assigned to one of three groups: a group that received a free one-month subscription to the* Post, *a group that received a free one-month subscription to the* Times, *and a group that received neither offer. Prior to the randomization, the sample was stratified into groups based on who they planned to vote for, whether they subscribe to another*

---

[11] Initially half of the subjects were to be asked if they would like a free one-month subscription to a national newspaper as a thank you for completing the survey. But this question was dropped early in the study.

*(non-*Post*, non-*Times*) newspaper, whether they subscribe to news magazines, and whether they were asked if they wished they read the paper more (50% of the subjects were asked this question). The stratification was designed so that Gerber, Kaplan, and Bergan had a balance on these covariates across the groups and the proportion of subjects in the groups was constant across strata. This randomization took place in two waves to maximize the time that subjects received the newspapers. We discuss such stratification techniques in Chapter 6. Households were given the option of canceling the subscriptions; approximately 6%, roughly equal between the* Post *and* Times*, canceled.*

The newspapers were unable to deliver to some of the addresses (76 of those assigned to the Times *and 1 of those assigned to the* Post*). Gerber, Kaplan, and Bergan employed a research assistant to monitor the delivery of newspapers to a random sample of households in the newspaper groups. While the* Post *had been delivered, the* Times *was not observed at all of the assigned addresses. Gerber, Kaplan, and Bergan spoke to the* Times *circulation department and called a small random sample of households assigned to receive the* Times *to verify their delivery.*

Moreover, when the lists of households to receive the newspapers were sent to the newspapers, 75 of the households assigned the Post *were reported to already subscribe to the* Post *(although it is unclear if they were subscribers only to the Sunday Edition or to the regular newspaper as well) and 5 of those assigned the* Times *were reported to already subscribe to the* Times*.*

After the gubernatorial election, Gerber, Kaplan, and Bergan reinterviewed 1,081 of the subjects, a response rate of approximately 32%. Gerber, Kaplan, and Bergan reported that (p. 11)

[t]he remainder was not reached because the individual refused to participate in the follow-up survey (29.7%), the individual asked for was not available at the time of the call (10.3%), the operator reached an answering machine (9.8%), or the individual only partially completed the survey (6%). The operators were unable to reach the remainder for a number of different reasons, including reaching a busy signal, being disconnected, or getting no answer on the phone. . . . The follow-up survey asked questions about the 2005 Virginia Gubernatorial election (e.g. did the subject vote, which candidate was voted for or preferred), national politics (e.g. favorability ratings for Bush, the Republicans, the Democrats, support for Supreme Court nominee Samuel Alito), and knowledge of news events (e.g. does subject know number of Iraq war dead, has subject heard of I. Lewis Libby).

Results: *Gerber, Kaplan, and Bergan found that those assigned to the* Post *group were eight percentage points more likely to vote for the Democratic candidate for governor than those not assigned a free newspaper. They also found similar evidence of differences in public opinion on specific issues and attitudes, but the evidence was weaker.*

Comments: *The results provide evidence that the biases in the news can affect voting behavior and political attitudes. The experiment is also a good illustration of randomization within strata and the measurement of causality using Intention-to-Treat, both of which we discuss in Chapter 5.*

Example 2.2 (Clientelism Field Experiment): *Wantchekon (2003) reported on a field experiment in Benin during a naturally occurring election in which candidates manipulated their campaign messages to test voter responses to messages of clientelist versus public policy messages.*

Target Population and Sample: *With the help of consultants, Wantchekon approached the leaders of four of the six political parties in Benin, which included the candidates of the two major parties. In Benin, voters are divided into eighty-four electoral districts. Wantchekon chose eight districts that are noncompetitive: four dominated by the incumbent government and four dominated by the opposition government. He also chose two competitive districts. The selection of districts was done in consultation with the campaign managers of the candidates. Within these districts, Wantchekon drew random samples for his postelection survey using standard survey sampling methods.*

Environment: *Benin is considered one of the most successful cases of democratization in Africa with a tradition of political experimentation. The election was a first-round election in which all expected a subsequent run-off election between the two major parties' candidates. Finally, candidates typically use a mixture of clientelism and public policy appeals in their election campaigns.*

Procedures: *Within each experimental district, two villages were chosen. In noncompetitive districts, one village was exposed to a clientelist platform, and the other a public policy platform. In the competitive districts, the manipulation differed; in one village one candidate espoused a clientelist platform and the other candidate a public policy platform, and in the other village the roles were reversed. The remaining villages in the selected districts were not exposed to the manipulation. The noncompetitive districts were ethnically homogenous and were less likely to be exposed to the nonexperimental manipulated campaign. The villages within each district were similar in demographic characteristics. Wantchekon took care to select villages that were physically distant from each other and separated by other villages so that, given normal communications, the manipulation was contained.*

The experimental platforms were carefully designed in collaboration with the campaign managers. The public policy message emphasized "national unity and peace, eradicating corruption, alleviating poverty, developing agriculture and industry, protecting the rights of women and children, developing rural credit, providing access to the judicial system, protecting the environment, and fostering educational reforms." The clientelism message consisted of a specific

promise to the village for things like government patronage jobs or local public goods, "such as establishing a new local university or providing financial support for local fishermen or cotton producers."

After devising the platforms, ten teams of campaign workers were created and trained. Each team had two members, one a party activist and the other a nonpartisan research assistant. The team trained, monitored, and supervised campaign workers. There were also statisticians who served as consultants. Wantchekon (p. 410) describes how the messages were conveyed to voters:

During each week for three months before the election, the campaign workers (one party activist and one social scientist) contacted voters in their assigned villages. With the help of the local party leader, they first settled in the village, contacted the local administration, religious or traditional authorities, and other local political actors. They then contacted individuals known to be influential public figures at home to present their campaign messages. They met groups of ten to fifty voters at sporting and cultural events. They also organized public meetings of fifty to one hundred people. On average, visits to household lasted half an hour and large public meetings about two hours.

In the post-election surveys, voters were asked demographic characteristics, degree of exposure of messages, and voting behavior.

Results: *Wantchekon found that clientelism worked as a campaign message for all types of candidates but was particularly effective for regional and incumbent candidates. He also found that women had a stronger preference for public goods messages than men.*

Comments: *Wantchekon's experiment is an unusual example of a case where political candidates were willing to manipulate their messages substantially in a naturally occurring election. The experiment raises some ethical issues of the influence of experimentalists in the DGP, although most of the manipulation took place in noncompetitive districts in an election that was widely seen as not significantly consequential given the likelihood that a run-off election would be held. We return to these ethical issues in Chapters 11 and 12.*

Example 2.3 (Negative Advertising Internet Survey Experiment): *Clinton and Lapinski (2004) report on an Internet survey experiment on the effects of negative advertising on voter turnout.*

Target Population and Sample: *Clinton and Lapinski used a national panel in the United States created by Knowledge Networks (KN). Information on KN can be found at http://www.knowledgenetworks.com/index3.html. Another Internet-based survey organization which has been used by political scientists is Harris Interactive (see http://www.harrisinteractive.com/). The*

panelists were randomly selected using list-assisted random-digit-dialing sampling techniques on a quarterly updated sample frame from the entire U.S. telephone population that fell within the Microsoft Web TV network, which at the time of the study was 87% of the U.S. population. The acceptance rate of KN's invitation to join the panel during the time of the study averaged 56%. Clinton and Lapinski randomly selected eligible voters from the KN panel for their study.

Subject Compensation: The panelists were given an interactive television device (Microsoft Web TV) and a free Internet connection in exchange for participating in the surveys. Participants are expected to complete one survey a week to maintain the service.

Environment: Clinton and Lapinski conducted their experiment during the 2000 presidential general election campaign and they used actual advertisements aired by the two major candidates, Bush and Gore. Subjects took part in the experiment in their own homes, although the subjects had to use the Web TV device to participate in the experiment. This somewhat reduced the variance in subjects' survey experience.

Procedures: An email was sent to the Web TV account of the selected subjects, informing them that their next survey was ready to be taken. Through a hyperlink, the subjects reached the survey. The response rate was 68% and on average subjects completed the survey within 2.7 days of being sent the email. The subjects were asked a variety of questions, both political and nonpolitical, for other clients of KN, but Clinton and Lapinski's questions were always asked first. During the survey, those subjects who had been randomly chosen to see one or more political advertisements were shown a full-screen advertisement and then asked a few follow-up questions.

The subjects were approached in two waves. The two waves in the experiment tested between different manipulations. In wave I, Clinton and Lapinski investigated the effect of being shown a single or pair of advertisements on Gore on the likelihood of voting, and in wave II, Clinton and Lapinski investigated the effect of seeing a positive or negative Bush advertisement conditioned on seeing a Gore negative advertisement. In wave I, subjects were divided into four groups depending on the types of advertisements shown: manipulation A (Gore negative and positive), manipulation B (Gore positive), manipulation C (Gore negative), and a group that was not shown an advertisement. Wave I took place between October 10, 2000, and November 7, 2000, with the median respondent completing his or her survey on October 12, 2000. In wave II, subjects were divided into three groups: manipulation D (Gore negative, Bush positive), manipulation E (Gore negative, Bush negative), and a group that was not shown an advertisement. Wave II took place between October 30,

2000, and November 5, 2000, and the median respondent completed his or her survey on November 1, 2000. Overall 2,850 subjects were assigned to groups A, B, and C; 2,500 were assigned to groups D and E. In wave I, 4,614 subjects did not see an advertisement; in wave II, 1,500 did not see an advertisement.

After being shown the ad or ads in both waves, subjects were asked the likelihood that they would vote in the presidential election. The question wording was slightly different in the two waves, with five options in wave I, and ten options in wave II. Finally, after the election, subjects were surveyed again, asking whether they had voted; 71% of the subjects responded to the post-election survey request.

Results: *Clinton and Lapinski found no evidence that the negative advertisements demobilize voters either using the initial probability of voting question or the post-election self-reported turnout question. They also found, when they controlled for respondent characteristics, that there is no mobilization effect of the campaign advertisements either. They argued that their results suggest that the effects of the manipulations are dependent on voter characteristics and the issues discussed in the advertisements and not the overall tone of the ads.*

Comments: *The group that did not see an advertisement in wave I was not a random sample devised by Clinton and Lapinski, but due to a technical difficulty that was known prior to the administration of the survey. However, Clinton and Lapinski state that the group was "essentially" random and that they used demographic controls in the analysis. Clinton and Lapinski analyzed the data using both the manipulations as independent variables and other demographic variables that can matter for turnout and that varied by manipulation group. We discuss the reasoning behind these estimation strategies in Section 4.2.8.*

Example 2.4 (Candidate Quality Lab Experiment): *Kulisheck and Mondak (1996), hereafter Mondak1; Canache et al. (2000), hereafter Mondak2; and Mondak and Huckfeldt (2006), hereafter Mondak3, reported on a series of experiments investigating how voters respond to information about the quality of candidates, independent of issue positions. Mondak et al. refers to all three experiments.*

Target Population and Sample: *Mondak et al. used as subjects undergraduate students enrolled in political science classes at universities in Mexico, Venezuela, and the United States. Mondak1 used 452 students at the University of Pittsburgh; Mondak2 used 130 students at two universities in Caracas (Univeridad Católica Andrés Bello and Universidad Simón Bolívar) and 155 students at three universities in Mexico (Universidad de las Américas, Universidad Autónoma de Méjico-Xochimilco, and Centro de Investigación y Docencia Económica de Méjico); and Mondak3 used 223 students at Indiana University.*

Subject Compensation: *Mondak et al. did not report whether the subjects were compensated for their participation. Presumably, however, they were compensated by credit in the political science classes they were taking.*

Environment: *The experiments reported on in Mondak1 and Mondak2 were conducted in classrooms using pen and paper. The experiments reported on in Mondak3 were conducted using computers. In all of the experiments, the candidates that subjects were presented with were hypothetical. In Mondak1 and Mondak2, subjects were given detailed information about the hypothetical candidates and asked to read material similar to what would appear in a local newspaper in a "meet the candidates" format. In Mondak3, subjects were presented information about the candidates via computer, but the information was more limited. An important factor in the experiments was that while in the United States there is usually much discussion about personal skill and the integrity of candidates for Congress, for legislative positions in Mexico and Venezuela the electoral system in place at the time of the experiments did not encourage voter discussion of these issues and voters rarely had this sort of information about the candidates.*

MediaLab and DirectRT are computer software programs designed for psychology experiments and used by political psychologists like Mondak et al. Information on this software can be found at http://www.empirisoft.com/ medialab.aspx and http://www.empirisoft.com/DirectRT.aspx.

Procedures: *First subjects took a survey about their political attitudes and attentiveness. Then subjects were presented with the information about the hypothetical candidates, either in paper form or via computer. All of the experiments varied the content of the information presented and subjects were randomly assigned to these manipulations. In all of the experiments, the researchers focused on manipulations of evaluations of the skill and integrity of the candidates. In Mondak1, the subjects were asked to give feeling thermometer-like ratings to the two candidates in each set and to identify which one would receive their vote. In Mondak2, the subjects were only asked which candidate they would vote for. And in Mondak3, subjects were asked whether they favored or opposed each candidate on an individual basis; that is, the candidates were presented not as pairs but singly on the computer screens. Also in Mondak3 the researchers measured the response time of subjects (how long it took for a subject to express his or her choice after seeing the information about a candidate).*

Results: *Mondak1 and Mondak2 found significant evidence that the qualities of the candidates affected the subjects' choices. They found this result was robust even when controlling for the importance of political issues for subjects and the distance between subjects' views on ideology and the candidates. Mondak3 found a similar effect but also found that subjects' attitudes toward candidates' competence and integrity were highly cognitively accessible*

(as measured by the response time). But they found no evidence that candidate character serves as a default basis for evaluation when things like partisanship and ideology are unavailable.

   Comments: *In Mondak3, the researchers also reported on survey evidence that supports their conclusions.*

Example 2.5 ("In-Your-Face" Discourse Lab Experiment): *Mutz   (2007) reported on laboratory experiments designed to evaluate the effects of tele-vised political discourse on awareness of opposing perspectives and views of their legitimacy.*

   Target Population and Sample: *At least 171 subjects were recruited from temporary employment agencies and community groups. Mutz did not report the community from which the subjects were drawn, although probably they were from the area around her university.*[12]

   Subject Compensation: *Subjects from the temporary employment agencies received an hourly rate that depended on whether the subjects came to campus to participate in this particular experiment or a set of studies over several hours. The subjects from civic groups participated as a fund-raising activity for their organizations.*

   Environment: *The experiments took place in a university facility where subjects were shown a 20-minute mock television program while sitting on a couch. The program was produced professionally with paid actors and a professional studio talkshow set was used to tape the program. The program was also professionally edited. The program was an informal political discussion between two "candidates" for an open congressional seat in a distant state, with a moderator who occasionally asked the candidates questions. Subjects were led to believe that the program and candidates were "real."*

   The candidates in the video had opposing views on eight different issues. The views drew on arguments from interest groups and the issues were topical at the time of the experiment. Four versions of the video were produced; two were civil versions and two were uncivil ones. In all four the same issue positions and arguments were expressed in the same words. As Mutz relates (p. 625):

The only departures from the script that were allowed for purposes of creating the variance in civility were nonverbal cues (such as rolling of the eyes) and phrases devoid of explicit political content (such as "You have completely missed the point here!"). The candidates in the uncivil condition also raised their voices and interrupted one another. In the civil version, the politicians spoke calmly throughout and were patient and respectful while the other person spoke.

---

[12] Mutz did not report the number of subjects who participated in experiment 3, so this sum only includes the participants in experiments 1 and 2.

Mutz did a manipulation check with pretest subjects who rated the candidates on measures of civility.

Mutz also manipulated the camera perspective. That is, in one of the civil versions and one of the uncivil versions there was an initial long camera shot that showed the set and location of the candidates and moderator, and then the subsequent shots were almost exclusively tight close-ups. In contrast, in the medium version the candidates' upper bodies were shown.

General Procedures: *After giving consent to participate, subjects were seated on the couch and given a pretest questionnaire. They were then asked to watch the video program and informed that they would be asked some questions after the program was concluded. Afterward, a paper-and-pencil questionnaire was administered. Subjects only saw four issues discussed, which varied by experimental session. In the questionnaire, Mutz asked open-ended questions designed to measure the extent that subjects recalled the arguments and the legitimacy of the arguments. Mutz also asked subjects to rate the candidates on a feeling thermometer. Using these basic procedures, Mutz conducted three different experiments.*

Experiment 1 Procedures: *In the first experiment, which used 16 subjects, the subjects saw a discussion using all four different combinations of camera perspective and civility (the issues also varied). These subjects' arousal during the video was measured using skin conductance levels by attaching two electrodes to the palm of each subject's nondominant hand. According to Mutz (p. 626), "Data collection began at the start of each presentation, with a 10-second period of baseline data recorded while the screen was blank prior to the start of each debate."*

Experiment 2 Procedures: *The second experiment used 155 subjects and the subjects saw only one of the four possible experimental manipulations. Subjects were randomly assigned to manipulations. Also included was a group of subjects who were randomly assigned to watch a nonpolitical program for the same amount of time and received the same questionnaire.*

Experiment 3 Procedures: *Mutz did not report the number of subjects used in this experiment. Experiment 3 is a partial replication of experiment 2 with one exception: in this experiment, all subjects saw the close-up versions of the videos and were randomly assigned to either civil or uncivil discourse.*

Results: *In experiment 1, Mutz found that uncivil discourse was significantly more arousing than civil discourse and that the close-up camera perspective was also significantly more arousing than the medium perspective. In experiment 2, she found that the awareness of rationales for arguments was also affected by the manipulations in the same direction, and uncivil close-up conditions led to the most recall. Furthermore, she found that the difference in thermometer ratings between the subjects' preferred and nonpreferred candidates*

*was not affected by civility in the medium camera perspective. However, in
the close-up camera condition, this difference was significantly greater in the
uncivil condition. The effect worked in both directions; that is, in the civil close-
up condition the difference in ratings fell and in the uncivil close-up condition
the difference rose, in comparison to the medium camera condition. Mutz
found a similar relationship in the perceived legitimacy of opposing arguments
in both experiment 2 and experiment 3.*

*Comments: Mutz's experiments are a good example of how control can be
exercised over unobservable variables in experiment 1, which we discuss further
in the next two chapters. Experiment 3 is an interesting instance of a researcher
replicating a previously found result, something we discuss in Chapter 7.*

Example 2.6 (Swing Voter's Curse or SVC Lab Experiment): *Battaglini
et al. (2008, 2010) report on a series of experiments conducted to evaluate the
predictions from the Swing Voter's Curse.*

*Target Population and Sample: Battaglini, Morton, and Palfrey recruited
student volunteers at Princeton University (84 subjects) and New York Univer-
sity (NYU; 80 subjects) from existing subject pools which had been recruited
across each campus. No subject participated in more than one session. The
subject pool for the experiments had been recruited via email to sign up for
experiments conducted at either the Princeton Laboratory for Experimental
Social Sciences at Princeton or the Center for Experimental Social Sciences
at NYU. One free online recruitment system is ORSEE (Online Recruitment
System for Economic Experiments, devised by Ben Greiner at the University of
New South Wales; see ttp://www.orsee.org/).*

*As is typical in political economy laboratory experiments, more subjects than
the required number were recruited because the experiments were designed for
specific numbers of participants. Subjects were chosen to participate on a first-
come/first-serve basis, and subjects who arrived after the required number of
participants had been met were given the show-up fee as payment.*

*Subject Compensation: Subjects were paid in cash based on their choices
during the experiment as described in the procedures. Average earnings were
approximately $20. In addition, subjects were also given a show-up fee of $10.
Subjects were assigned experiment-specific identification (ID) numbers and
payments were made to subjects by ID numbers such that records were not kept
that could be used to match subject identity with payments received or choices
in the experiment.*

*Environment: The experiments used a standard setup for computerized
laboratory experiments by political economists. That is, the experiments
were conducted in computer laboratories via computer terminals and all*

communication between the experimenter and subjects was conducted via the computer interface. Each subject's computer screen was shielded from the view of other subjects in the room through privacy screens and dividers. Subjects were first presented with instructions about the experiment and then took a short quiz regarding the information in the instructions before they were allowed to continue to the experiment. Subjects were told all the parameters of the experiment as described in the procedures. The experimental parameters were chosen specifically to evaluate game-theoretic predictions and a formal model was used to derive these predictions explicitly. The software used for the experimental program was multistage, which is an open-source software program for laboratory experiments developed at the California Institute of Technology (see http://multistage.ssel.caltech.edu/). Of particular usefulness for experiments is the free software z-Tree for the Zurich Toolbox for Readymade Economic Experiments (see http://www.iew.uzh.ch/ztree/index.php), which was devloped by Urs Fischacher.

***Procedures:** Before the experiment began, one subject was randomly chosen to be a monitor. The monitor was paid a flat fee of $20 in addition to his or her show-up fee. Each session was divided into periods. In five of the sessions conducted at Princeton University, 14 subjects were randomly assigned to two groups of 7 voters. The group assignments were anonymous; that is, subjects did not know which of the other subjects were in their voting group. In two sessions at Princeton, only 7 subjects were in each session, so in each period the group of voters was the same. At NYU all subjects in each session were in the same group; three sessions used 21 subjects and one session used 17 subjects.*

In each period and group, the monitor would throw a die to select one of two jars, red or yellow. Although subjects could see the monitor making the selection, they were not able to see the selection made. The red jar contained two red balls and six white balls, the yellow jar contained two yellow balls and six white balls. These were not physical jars, but were represented on the computer monitors. Subjects then were shown the jar with eight clear balls, that is, a jar without the colors. They then could click on one of the balls and the color of the ball selected would be revealed. If a red or yellow ball was revealed, they learned which jar had been chosen. If a white ball was revealed, they did not learn which jar had been chosen.

After choosing a ball and finding out its color, the group members simultaneously chose where to abstain, vote for red, or vote for yellow. The computer casts a set number of votes for the red jar. The jar that received the majority of the votes was declared the winner, including the computer votes (ties were broken randomly by the computer). If the jar chosen by the majority was the correct jar, all the subjects earned a payoff of 80 cents in the period, and if the

jar chosen by the majority was the incorrect jar, all the subjects earned a payoff of 5 cents in the period.

Subjects were told the outcome of the period in their group. They were then randomly reassigned to new groups for the next period if applicable and the procedure was repeated. The colors of the balls within each jar were randomly shuffled each period so that if a subject repeatedly chose to click on the same ball, whether they were revealed a white ball was randomly determined by the percentage of white balls in the jar (i.e., the probability of observing a white ball was always 75%). A check of the data shows that the procedure worked as desired: approximately 75% of subjects saw a white ball and 25% saw either a red or yellow ball. There were a total of 30 periods in each session.

Each session was divided into three subsessions of ten periods each. In one of the subsessions, the computer had zero votes. In sessions with groups of 7 voters, in one subsession the computer had two votes and in one subsession the computer had four votes. In sessions with groups of 17 and 21 voters, in one subsession the computer had six votes and in one subsession the computer had twelve votes. The sequence of the subsessions varied by session. The following sequences were used depending on the number of voters in the groups: (0,2,4), (0,4,2), (2,4,0), (4,0,2), (4,2,0), (0,6,12), and (12,6,0).

Battaglini, Morton, and Palfrey also varied the probability by which the monitor would pick the red jar. In some sessions, the probability of a red jar being selected was equal to the probability of a yellow jar, which was equal to 1/2. This was done by having the monitor use a six-sided die; if 1, 2, or 3 were shown, the red jar was selected, and if 4, 5, or 6 were shown, the yellow jar was selected. In other sessions the probability of a red jar being selected was equal to 5/9 while the probability of a yellow jar was equal to 4/9. This was done by having the monitor use a ten-sided die; if 1, 2, 3, 4, or 5 were shown, the red jar was selected; if 6, 7, 8, or 9 were shown, the yellow jar was selected; and if 10 was shown, the die was tossed again.

Results: Battaglini, Morton, and Palfrey found that subjects who were revealed either a red or yellow ball voted for the red or yellow jar, respectively. They also found that when the number of computer votes was equal to zero, most of the subjects who were revealed a white ball abstained. As the number of computer voters increased, uninformed voters increased their probability of voting for the yellow jar. These results occurred even when the probability of the red jar was 5/9.

Comments: The results strongly support the SVC theoretical predictions. Uninformed voters are more likely to abstain than informed voters, but when there are partisans (as operationalized by the computer voters), the uninformed voters appear to vote to offset the partisans' votes, even when the probability

is higher that the true jar is the red jar (the partisans' favorite). Battaglini, Morton, and Palfrey also considered some alternative theoretical models to explain some of the subjects' errors, an analysis that we explore in Chapter 6.

Example 2.7 (Polls and Information Lab Experiment): *Dasgupta and Williams (2002) reported on a laboratory experiment that investigates the hypothesis that uninformed voters can effectively use cues from public opinion polls as an information source to vote for the candidate they would choose if informed.*

Target Population and Sample: Dasgupta and Williams recruited 119 undergraduate student volunteers at Michigan State University. The authors recruited subjects who were unaccustomed to psychological experiments and were unfamiliar with spatial voting models and formal decision theory.

Subject Compensation: Subjects were paid based on their choices as described in the procedures. As in many political economic experiments, the subjects' payoffs were denominated in an experimental currency that was then converted at the end of the experiment into cash. Dasgupta and Williams called this experimental currency "francs." We discuss reasons for using an experimental currency in Section 10.1.4; one reason for doing so in the Dasgupta and Williams experiment is that it allowed them to have different exchange rates for different types of subjects, as described later. The exchange rates were fixed and known by subjects. Subjects earned on average $22 for the two-and-a-half hours plus a show-up fee.

Environment: As in Example 2.6, the experiment was conducted via a computer network. Subjects were seated so that they were unable to see the computer monitors and choices of other subjects. The experimental parameters were chosen to fit the formal model presented by Dasgupta and Williams.

Procedures: Dasgupta and Williams conducted six experimental sessions with 17 subjects in each session. At the beginning of a session, two subjects were chosen to be incumbent candidates; the remaining subjects were assigned as voters. The sessions were divided into two subsessions lasting 10 periods each. The incumbent candidates were randomly assigned to separate subsessions and only participated in the experiment in the subsession to which they were assigned. At the beginning of each subsession, the incumbent candidate was assigned an issue position of either 0 or 1000, which was held fixed throughout the subsession. This issue position was publicly announced to all subjects. In each period in a subsession, the subjects were randomly divided into three equal-sized groups of 5 voters each with issue positions of 250, 500, and 750, respectively. So in each period, a subject's issue position was a new random draw. All subjects knew the distribution of issue positions of voters.

Table 2.1. *Voters' payoffs*

Incumbent	Incumbent's issue position = 0					
	Voter issue positions					
	Incumbent wins			Challenger wins		
Quality	250	500	750	250	500	750
10	40	10	10	25	10	20
20	45	45	25	30	45	40
30	50	50	50	35	50	60

Before the subsession began, the incumbent candidate was provided with an initial endowment of 900 francs. The candidate chose an "effort" level equal to either 10 or 20. Effort levels were costly to the candidate; if he or she chose an effort of 10, the cost was 30 francs. The cost of an effort of 20 was 170 in one subsession and 90 in the other. The cost of the effort chosen by the candidate was deducted from his or her endowment. The computer program then assigned the candidate a "quality" of either 10, 20, or 30, with equal probability. Note that the incumbent candidate was not told his or her quality before choosing an effort level. The effort level and quality were combined to produce an "output." Three voters in each of the voting groups were randomly chosen to be told the output. The remaining two voters in each voting group received no information about the output and were uninformed.

Voters then participated in an election. They could either vote for the incumbent or the challenger (who was an artificial actor). If the incumbent won the election, he or she would receive 300 francs. The voters' payoffs depended on their own issue position, the issue position of the incumbent candidate, and the quality of the incumbent. Table 2.1 presents the payoffs to voters by issue position when the incumbent candidate's issue position equals 0 (the case where the incumbent candidate's issue position equals 1000).

Before the election was held, three separate opinion polls were taken. In each poll, voters were asked which of the two candidates – incumbent or challenger – he or she currently preferred. All subjects were revealed the aggregate outcomes of the polls. After the polls were completed, the election was held, which was decided by majority rule (ties were broken randomly by the computer). The incumbent candidate and the voters received their respective payoffs and a new period began with voters reassigned to new voter types.

Dasgupta and Williams also conducted an additional seventh session which was the same as the other sessions except that the candidate's issue position was held constant at 0 in both subsessions and all voters were uninformed.

Results: *Dasgupta and Williams found significant support for the theoretical argument that uninformed voters will use poll information to make more accurate voting decisions and that the electorate's behavior is consistent with complete information. Compared to the session in which no voters were informed, voters were much more likely to choose optimally. However, voters make errors when incumbent candidates choose suboptimally. Specifically, in the low-cost subsessions, according to the theory, the candidates should always choose a low level of effort and in the high-cost trials the candidates should choose a high level of effort. The majority of the time candidates chose effort levels as predicted, with errors occurring early in the subsessions. Voters' mistakes were most likely to occur when candidates did not choose effort levels as predicted.*

Comments: *The experiments are an extension of earlier experiments and theoretical work reported by McKelvey and Ordeshook (1984, 1985a,b, 1986a,b), and Collier et al. (1987) on a rational expectations model of the relationship between voting and information. Dasgupta and Williams also compared the predictions of several alternative models in which uninformed voters are uninfluenced by polls and found support for the rational expectations model.*

Example 2.8 (Decentralization Experience and Turnout Natural Experiment): *Lassen (2005) presents a study of the effects of information on voter turnout in Copenhagen. He exploits a naturally occurring situation that affected voter information in a naturally occurring referendum and studies the effect of the information on turnout in the referendum.*

Target Population and Sample: *Lassen used a telephone survey of voters in the city of Copenhagen. The voters in the city were partitioned into five strata reflecting five different areas of the city. In Denmark every citizen has an identification number. Everyone eligible to vote automatically receives a ballot card at their address in the census registry. The voter then presents the ballot card at their local polling station and if the individual votes, it is registered. The survey was commissioned by the city, so the sample was drawn from the city's list of eligible voters in the election. The response rate to the survey was 55%, resulting in a sample of 3,021 observations. Of this sample, one-third refused to answer questions about their yearly income or whether they had voted in the referendum, which left 2,026 observations. Lassen used the larger sample for some analysis as well as the smaller sample.*

Environment: *In 1996 the city of Copenhagen decided to conduct a policy experiment in decentralization of city government by having some areas of the city experience decentralized services while other areas continued to experience centralized services. The policy experiment lasted for four years and in 2000 a*

citywide consultatory referendum was held on whether the program should be extended to or abolished in the entire city.[13]

Procedures: *The city was divided into fifteen districts (divisions which did not exist prior to the policy experiment): eleven districts where government services continued to be centrally administered and four districts where approximately 80% of government services were administered by locally elected district councils, which are labeled pilot city districts or PCDs. The designers of the policy experiment attempted to choose four districts that were representative of the city. The strata for the survey were the four PCDs and the rest of the city.*

The consultatory referendum was held on the same day as a nationwide referendum on whether Denmark should join the common European currency. It was possible for individuals to vote in only one of the referenda if they wished.

The survey asked respondents if they voted in the last municipal election, whether they voted in the nationwide referendum, whether they voted in the municipal referendum, their opinion of the decentralization experiment, their opinion of the responsiveness of municipal council members, how interested they are in political issues, and a set of demographic questions.

Results: *Lassen found that turnout was significantly higher among informed voters than in the districts where services remained centralized. The effect is robust to a number of different specifications and is strongest among those voters who had zero cost of voting; that is, those who voted in the nationwide referendum and so had paid the cost of going to the polling place.*

Comments: *Lassen dealt with a wide variety of methodological issues in identifying and estimating the causal relationship between information and voting. We discuss his study more expansively as we explore these issues. As in the other experiments we have so far discussed, in Lassen's study the manipulated variable is not the same as the treatment variable. The manipulated variable is whether the respondent lived in the district that experienced the decentralization policy experiment, whereas the treatment variable is the information level of the respondents. As a proxy for the treatment variable, Lassen uses whether a respondent reported an opinion of the decentralization experiment, classifying those with no opinions as uninformed. He cites empirical evidence that finds a strong correlation between other measures of voter information and the willingness to express an opinion. Lassen also finds, however, that this treatment variable is endogenous and affected by other observable variables from the survey. We discuss how Lassen deals with this endogeneity in identifying and estimating the causal effects of information on turnout in Chapter 5.*

---

[13] The referendum was nonbinding because the Danish constitution does not allow for binding referenda at the municipal level.