Experimental Political Science and the Study of Causality

From Nature to the Lab

Rebecca B. Morton
Kenneth C. Williams

CAMBRIDGE

2

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.*

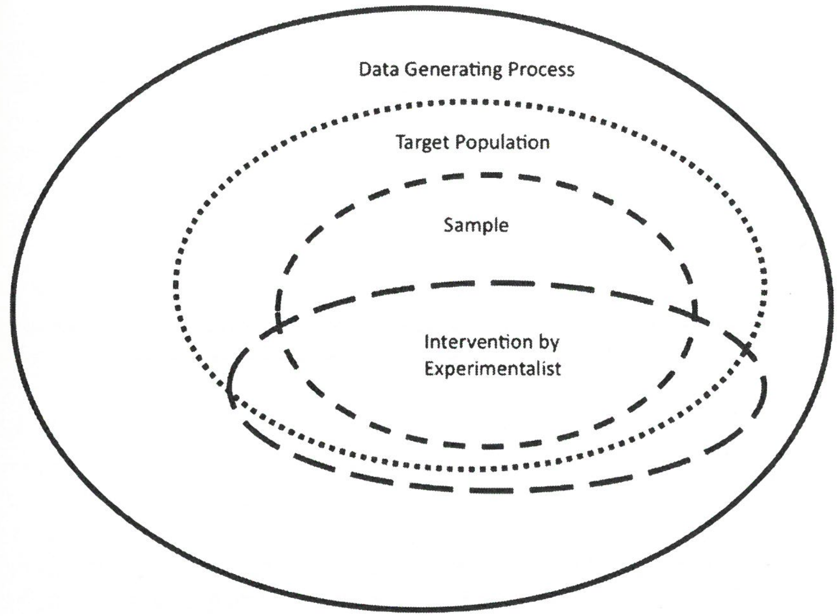

Figure 2.1 Relationships between the Data Generating Process, Target Population, Sample for Study, and Experimental Intervention.

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