CHAPTER 1

The Challenge of Inference

CONTENTS

There is no doubt that political observers believe that there has been a fundamental change in the nature of U.S. politics over the past few decades. David Broder noted in 2006 that "the terms 'gridlock' and 'polarization' have become staples of the political vocabulary."[1] Ezra Klein wrote in 2012, "We use 'polarization' as an epithet. It's what's wrong with America's politics. It's what's wrong with America's political parties. It's what's wrong with America's politicians. It's what's wrong, finally, with America."[2] But to what extent has America actually become more polarized in recent years? You may be surprised to learn that there is actually some debate over this question among political scientists. Stanford University professor Morris Fiorina wrote a book shortly after the 2004 presidential election posing the question, "Culture War?" Fiorina takes issue with the notion that the public has become more ideologically polarized over the past several decades, arguing that this is actually an elite phenomenon that is not duplicated at the mass level:

> Americans are closely divided, but we are not deeply divided, and we are closely divided because many of us are ambivalent and uncertain, and

1 David Broder, "Behind the Gridlock," *Washington Post*, November 2, 2006, p. A17.

2 Ezra Klein, "Olympia Snowe is right about American politics. Will we listen?" *Washington Post*, February 28, 2012. Accessed July 4, 2013. www.washingtonpost.com/blogs/wonkblog/post/olympia-snowe-is-right-about-american-politics-will-we-listen/2011/08/25/gIQA3KkwgR_blog.html.

> consequently reluctant to make firm commitments to parties, politicians, or policies. We divide evenly in elections or sit them out entirely because we instinctively seek the center while the parties and candidates hang out on the extremes.[3]

According to Fiorina, Americans are mostly moderate, caught somewhere in between the polarized parties. Fiorina's view is contested by other political scientists, however. Alan Abramowitz and Kyle Saunders find evidence that the most politically engaged and active citizens take very polarizing positions.[4] In essence, they argue that those who identify with a political party have become more polarized in recent decades:

> There are sharp divisions between supporters of the two major parties that extend far beyond a narrow sliver of elected officials and activists. Red state voters and blue state voters differ fairly dramatically in their social characteristics and political beliefs. Perhaps most importantly, there is a growing political divide in the United States between religious and secular voters. These divisions are not the result of artificial boundaries constructed by political elites in search of electoral security. They reflect fundamental changes in American society and politics that have been developing for decades and are likely to continue for the foreseeable future.

In other words, the existing research points us in two different directions.

The debate over polarization is by no means the only thing that political scientists argue about. Indeed, for almost every research study produced by a political scientist, there are at least some scholars who question the validity of that study's findings. The fact that political scientists engage in spirited debates is testament to how challenging it is for us to generate knowledge in the first place and how seriously we take the enterprise.

In this chapter, we introduce some of the key reasons why making inferences, even those as basic as whether or not the public is polarized, can be so challenging for political scientists. By calling attention in this first chapter to the challenges a researcher faces when attempting to make sound inferences, we can then use subsequent chapters to help you understand how to overcome those challenges. We begin with some examples of the challenges that political scientists have faced in answering a diverse set of research questions.

3 Morris P. Fiorina, Samuel J. Abrams, and Jeremy Pope, *Culture war?* (New York: Pearson Longman, 2005).

4 Alan Abramowitz and Kyle Saunders, "Why can't we all just get along? The reality of a polarized America," *Forum* 3 (2005): 1–22.

THREE ADDITIONAL EXAMPLES

Does Campaign Advertising Work?

Early attempts to answer this question found that campaign advertising did not seem to have much impact. From the 1940s through the 1980s, most studies on American elections relied heavily on survey research. Public opinion surveys were invaluable for studying how Americans thought about politics and what influenced them to vote the way they did. However, public opinion surveys were quite limited as an instrument for studying the effects of campaign advertising. During this period, scholars would typically test whether campaign advertisements had influenced vote decisions by asking individuals whether they recalled seeing any of the candidates' advertisements during the campaign and then seeing whether those who could recall seeing advertisements voted differently than those who did not see any.

This approach faced a significant inferential challenge, however. To measure exposure to campaign advertising, studies during this period relied on survey questions asking individuals whether they recalled seeing advertisements aired by the candidates. Yet, *seeing* an advertisement and *recalling* that you saw an advertisement are by no means equivalent. In fact, one set of experiments run by two political scientists discovered that about half of all individuals who had been exposed to a campaign advertisement did not recall having seen that advertisement just thirty minutes later.[5] Furthermore, individuals who recall seeing advertisements tend to be more interested in politics, more supportive of a particular party, and more loyal to a particular candidate.[6] These are precisely the types of individuals who are the least likely to be susceptible to advertising effects.

Thus, studies of advertising effects that relied exclusively on survey data appear to have understated the effects of campaign advertisements since the types of voters who fail to recall seeing political advertisements are the ones who are most likely to be persuaded by those advertisements. The problem confronting these early studies was a measurement problem. In recent decades, scholars have turned to new techniques and approaches that have overcome some of these inferential challenges and generated different conclusions about the influence of campaign advertising. We describe these techniques in much more detail in Chapters 5 and 6.

5 Stephen Ansolabehere and Shanto Iyengar, *Going negative* (New York: Free Press, 1995).

6 Shanto Iyengar and Adam F. Simon, "New perspectives and evidence on political communication and campaign effects," *Annual Review of Psychology* 51 (2000): 149–169.

Do Ethnic Divisions Cause Civil Wars?

Conventional wisdom posits that civil wars are the product of ethnic and religious division, and early scholarship echoed this sentiment. One could point to salient cases such as the Holocaust in Europe, conflict between the Hutus and the Tutsis in Rwanda and Central Africa, and civil war in the former Yugoslavia between Serbs, Bosnians, Croatians, and other groups. Based on these prominent cases, it is tempting to conclude that ethnic divisions are the key drivers of conflict.

As data on civil wars, conflict, and ethnic and religious division have developed and improved, however, more recent scholarship has been better able to test this relationship. By comparing civil wars from 1945 to 1999, for example, Fearon and Laitin find that more ethnically or religiously diverse countries are no more likely to experience civil war.[7] These findings have been echoed by other scholars, including Bates, Collier, and Hoeffler.[8] While the instances of ethnic conflict mentioned above are so prominent in our collective conscience, there are far more civil wars that are driven by conflicts over resources or opportunistic political entrepreneurs. Furthermore, there are many more diverse societies that are entirely peaceful. In the language used in the discussion that follows, previous studies of ethnicity and conflict suffered from sampling, or case selection, problems—they studied cases that were more likely to be driven by a theory that ethnic divisions caused civil wars.

Is Privatization or Government Intervention Necessary to "Govern the Commons?"

In his influential article on governance of common-pool resources (e.g. pastures, water resources, forests, fisheries, irrigation systems), Garrett Hardin coined the term "the tragedy of the commons."[9] He described a pasture open to all and observed that an individual herder gains the benefits of his own animals grazing and shares the cost of overgrazing on the pasture with everyone else who is using the commons. He noted that it is in the herder's short-term, best interest to increase his own herd and absorb more of the benefits of the pasture while only bearing a share of the costs. If all herders increase their own herd and capture the benefits of the pasture, however, the pasture will quickly become overgrazed and usable to no one, hence the tragedy of the commons. Subsequent authors proposed two solutions to governing common-pool

7 James D. Fearon and David D. Laitin, "Ethnicity, insurgency, and civil war," *American Political Science Review* 97 (2003): 75–90.

8 Robert H. Bates, *When things fell apart* (Cambridge, UK: Cambridge University Press, 2008); Paul Collier, Anke Hoeffler, and Catherine Pattillo, "Flight capital as a portfolio choice," *World Bank Economic Review* 15 (2001): 55–80.

9 Garrett Hardin, "The tragedy of the commons," *Science* 162 (1968): 1243–1248.

resources: either the government could take over the commons and manage it, or the land could be divided up and privatized.

In her influential work on governing the commons, however, Elinor Ostrom argues that Hardin's theory and the policy prescriptions that emanated from it were not empirically tested.[10] For example, subsequent research revealed that nationalization of land in order to protect it often led to disastrous consequences. To provide a more accurate understanding of how the commons are and should be governed, based on social goals, Ostrom and her colleagues collected and coded thousands of case studies on different common-pool resources throughout the world upon which to draw more accurate inferences. Surprisingly, Ostrom found that, in many cases, users of the resource were able to avoid the tragedy of the commons without privatization or government nationalization. This work was so important and path breaking that it ultimately earned Ostrom the Nobel Prize.

SOME BASIC TERMINOLOGY

As these examples suggest, making valid and reliable inferences is a challenge for political scientists. In this chapter, we introduce some of the more common challenges or threats to inference; however, it would be "challenging" to go much further in this book without first introducing some of the most basic terminology that political scientists employ. As we introduce this terminology, keep in mind that at such an early stage in the book it is difficult to fully contextualize this information; your understanding of these concepts will increase as we continue our discussion here and in subsequent chapters.

Nearly every research project starts with a **research question**. However, generating a good research question is actually quite challenging, as scholars and students are often tempted to study a wide range of questions rather than focus on a specific one. Furthermore, it is common for research questions to evolve or even change quite dramatically as one proceeds with a research project. The formulation of the research question is so crucial to a successful research project that we devote the entirety of the next chapter (Chapter 2) to this enterprise. For now, it is sufficient to note that a research question generally identifies some (political) phenomenon we wish to understand. For example, the previous section lists several potential research questions. Does campaign advertising work? Do ethnic divisions cause civil wars? Is privatization or heavy government intervention necessary to "govern the commons?" We might also refer back to the introduction for a research question: Does a candidate's race affect his/her electoral success?

10 Elinor Ostrom, *Governing the commons: The evolution of institutions for collective action* (Cambridge, UK: Cambridge University Press, 1990).

After formulating a good research question, we generally move on to constructing a **theory**. Broadly speaking, a theory is an idea about how we think the world works. More narrowly, a theory is typically a discussion of what we expect to be the most likely answers to our research question and, especially important, *why* we expect these answers to be correct over other possibilities. For example, in considering whether the race of a candidate affects his/her electoral success, we might construct a theory that explains that racial prejudice still exists among some white voters and that this prejudice would cause them to refuse to support a candidate of color whose political ideology they might have otherwise supported.

Theory building is also an essential part of the research enterprise because it provides significant guidance to how we should construct our empirical tests and which possible alternative explanations need to be ruled out. Despite the centrality of theory in every aspect of a research project, theory building is another aspect of the research process that students find challenging. Chapter 3 focuses directly on best practices for building theory in a research project and the remainder of this chapter will often refer back to the importance of this step for making strong inferences about the world.

Using our theory we can derive testable **hypotheses**. If a theory is a broad discussion of how the world works, a hypothesis is a specific statement based on our theory that we can test in the real world. For example, our theory above links racial prejudice to vote choice. From this theory we could derive a hypothesis, such as:

> *White voters will be less likely to vote for African American candidates than white candidates.*

Note that hypotheses are specific and testable statements, while theories are more general. Unlike a theory, a hypothesis does not explain why a researcher might expect a relationship to exist; it merely states what relationship is expected.

Typically, hypotheses identify at least one **dependent variable**, or a phenomenon that we want to explain, and at least one **independent variable**, a factor that we think does the explaining. In the example above, the vote choices of white voters is the dependent variable and the race of the candidate is the independent variable.

We use the term **variable** to denote the measurement or **operationalization** of concepts we are interested in studying. For example, we may be interested in gauging electoral support for African American electoral candidates, but this could be measured in a number of different ways, depending on the question. One way of measuring this is to estimate the percentage of the white vote won by all white candidates for a particular office compared with the percentage of the white vote won by all African American candidates.

However, there are other ways to measure support as well. For example, instead of counting votes, perhaps we have access to a public opinion poll

that asked respondents whether they had a favorable or unfavorable opinion of a number of different candidates. If we chose this approach, then we would want to adjust our hypothesis to reflect this somewhat different dependent variable. Our hypothesis might now be altered to read:

> *White voters will evaluate African American candidates less favorably than they evaluate white candidates.*

Note that this hypothesis is similar to the first, but we have adjusted the wording to reflect the different dependent variable we would be using. In the first example, our dependent variable was the vote; in our second example, our dependent variable was how favorably voters evaluate candidates.

After the researcher specifies her hypothesis (or hypotheses), the next step is to design a research study that will allow her to test this hypothesis. There are two broad approaches that political scientists can take to studying the world—observational and experimental. An **observational study** is one that collects information on the independent and dependent variables as they exist in the natural state of the world. Observational studies can take many forms, including quantitative studies that analyze a larger number of cases (see Chapter 6) or qualitative studies with a small number of cases (see Chapter 7). An observational study hypothesizing that whites are less supportive of minority candidates might examine the percentage of whites who voted for minority candidates for Congress compared with the percentage who voted for white candidates in the past several election cycles. If the study found that white voters tended to vote at higher rates for white candidates than minority candidates, then the evidence would support the hypothesis.

Experimental studies are different from observational studies in that we do not merely examine the independent variable as it exists in the world; rather, as researchers, we directly control the independent variable. Chapter 5 is dedicated to explaining the experimental approach. To offer an example: suppose we took a sample of white voters and asked them to choose between two fictitious candidates. We might describe the candidates in exactly the same way for each participant in our study, but for half of our respondents we might tell them that the candidate is an African American while for the other half we would describe that same candidate as being white. If participants who were told the candidate was African American were less likely to support that candidate than the participants who were told he was white, then this would again support our hypothesis that race influences support for African American candidates.

Regardless of the design chosen, we would never be able to study every instance in which a white voter casts a ballot for a minority candidate. As a result, we will never be able to definitively prove our hypothesis. We can, however, make inferences.

WHAT IS AN INFERENCE?

Gary King, Robert Keohane, and Sidney Verba define **inference** as "the process of using the facts we know to learn about facts we do not know."[11] This definition underscores the basic point that we can never know all of the facts, so the knowledge we construct is always built on inference. On a daily basis, consciously or subconsciously, we are constantly observing the world around us and making inferences based on these observations. We cannot directly know how someone is feeling, but we observe facial features, like a frown or a smile, and make an inference about those feelings. We may not have time to stop and talk to protesters downtown, but we may be able to infer from their signs or chants what is motivating them to protest. We may see a long line outside of a local restaurant and infer that they must serve delicious food.

There are two types of inferences we might formulate about the world—descriptive and causal. An easy way to think about the difference between these two types of inference is to consider the difference between describing something and explaining something. A **descriptive inference** is an inference we make about how the world is (or was)—it is the act of describing some aspect of the world. For example, you may notice that people you know seem to have more intense political disagreements than they used to. Based on this observation, you might infer that Americans are more politically polarized than they used to be. In this case, you would be using the facts you know (the intensity of political disagreements among your acquaintances) to infer something you cannot directly observe (how much Americans disagree about political issues): a descriptive inference.

In many cases, we want to go a step further than merely making a descriptive inference. In addition to knowing something about how the world is, we often want to know *why* the world is that way—we want to explain. For example, if we determine that Americans are more polarized than they used to be, we might want to know *why* they are more polarized. Answering this question will require us to make a different type of inference called a **causal inference**. Causal inferences are inferences we make about why something happens. This is where our theory and hypotheses come into play. If we conduct a study of racial attitudes and vote choice and conclude that racially prejudiced Caucasian voters are less likely to vote for Barack Obama, then we would be making a causal inference.

Descriptive and causal inferences are inherently related. Indeed, it is impossible to make a causal inference without first making a descriptive inference. After all, how can you know *why* Americans are more polarized without first

11 Gary King, Robert O. Keohane, and Sidney Verba, *Designing social inquiry: Scientific inference in qualitative research* (Princeton: Princeton University Press, 1994), p. 46.

knowing that they are in fact more polarized? Most frequently, a descriptive inference is what sparks our research question in the first place. For example, despite Fiorina's arguments against the notion that Americans are more polarized today, observations and evidence to the contrary have motivated dozens of political scientists to ask why this shift occurred or how polarization has influenced American politics and policymaking. To offer another example, several decades ago political scientists observed that nations with democratic governments rarely went to war with each other; this descriptive inference sparked an enormous body of scholarship attempting to explain what has been termed the "Democratic Peace."

THE CHALLENGE OF INFERENCE

Anybody can make inferences about the world—after all, we make inferences every day, often without realizing it. What is challenging is making *accurate*, or valid, inferences. Consider the example above: if we observed an increase in political arguments and inferred that people were more ideologically polarized than in the past, would this be a valid descriptive inference? We cannot know for sure whether our inference is correct; after all, if we knew the truth, we would not have needed to make an inference in the first place. But what we can do is evaluate how we arrived at that inference. The more defensible our method of making the inference, the more likely the inference is to be correct. In this section, we describe some of the common challenges we face when making descriptive and causal inferences. While we preview some of the solutions to those challenges here, the remainder of the book is largely reserved for exploring these challenges and potential solutions in greater detail.

Challenges to Descriptive Inference

It might seem like the process of making a descriptive inference would not be particularly challenging. But extending the example we discussed above will help to illuminate just how difficult the task can be. Here we summarize the inference we made about polarization and the data utilized to make that inference:

> *Data*: Witnessing more intense political arguments among our acquaintances.
> *Inference*: The American public is now more ideologically polarized.

On its face, the inference we drew seems perfectly reasonable in light of the data we have accumulated, but, as social scientists, we are trained to be

Box 1.1: How the Daily Stock Report Reveals the Differences between Journalists and Social Scientists

Most weekday evenings one can tune in to the news and hear a report about how the stock market fared that particular day. Reporters' stories on the stock market's performance almost always provide a straightforward accounting of whether the Dow Jones Industrial Average increased or decreased during the day's trading. But reporters typically do something else in these stories—they make causal inferences by attempting to explain why stocks rose or fell that day. Unfortunately, as political scientist Edward Tufte once wrote, "Explanations of daily changes in aggregate stock market indices are among the most ridiculous, speculative, and uncertain causal inferences made by journalists."[a]

The problem is that journalists generally are not using a systematic approach to making inferences about changes in the stock market. Instead, they first typically note that the stock market has changed, and then they search for something that can explain that change. Often, the causal stories they tell are plausible—perhaps the government released a report showing a decrease in unemployment and on that same day the Dow Jones increased by 100 points. The headline that evening is likely to note, "U.S. Jobs Report Gives Stocks a Lift" (as a February 3, 2012 *New York Times* headline read). But how do we know that the jobs report caused the stock market movement? Indeed, a substantial body of research produced by economists attempts to discern whether events like the releasing of employment reports causes stock market fluctuations. The association between the two is by no means clear. In fact, one study even found that, on average, the stock market performed *worse* when the government announced lower unemployment and *better* when the announcement noted higher unemployment.[b] Other studies have found no significant effect in either direction.[c]

Of course, journalists typically do not have time to carry out systematic research to support the inferences they make in their stories. Unfortunately, these reporters typically fail to convey how this lack of research affects how confident (or doubtful) their viewers should be about the inferences they are making on-air.

a. www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0000ml. Accessed August 22, 2013.
b. John H. Boyd, Jian Hu, and Ravi Jagannathan, "The stock market's reaction to unemployment news: Why bad news is usually good for stocks," *Journal of Finance* 60 (2005): 649–672.
c. Mark J. Flannery and Aris A. Protopapadakis, "Macroeconomic factors do influence aggregate stock returns," *Review of Financial Studies* 15 (2002): 751–782.

skeptics. A *social scientist should resist the urge to dismiss or accept an inference before evaluating how the inference was arrived at.*

Inferential challenges can come in many forms. First, we can arrive at an unsupported inference simply by failing to properly formulate or operationalize the concepts we are studying. In this case, our key concept is "ideological

polarization." How might we define the concept of ideological polarization? The best approach to this task is to draw on previous research. For example, Morris Fiorina and Samuel Abrams suggest: "Movement away from the center toward the extremes would seem to be a noncontroversial definition of polarizing."[12]

Concept definition is just the first step; once a concept is defined, a researcher must decide how to measure it. As discussed above, operationalization is the process of moving from the definition of a concept to determining how to measure that concept. In this case, we have operationalized polarization as the frequency with which we observe our acquaintances having intense political disagreements. Do you see the problem? Our operationalization is not an ideal measurement of our concept. While the frequency of political arguments may be related to political arguments, this is not the same thing as polarization as Fiorina and Abrams define it above. The former is an action and the latter an attitude. To be sure, an increase in political disagreements may be one observable implication of increased polarization, but an increase in political disagreements does not necessarily indicate more polarization.

Even if our measurement was a successful operationalization of our concept, we would face other challenges to making valid descriptive inferences. One such challenge comes in the form of **measurement error**—error we encounter when we actually go about measuring our concepts. In our example, measurement would occur every time we recognize and record an intense political argument among our sample. Consider the various sources of errors we might encounter in doing so. For example, we might miss some arguments or mis-categorize some friendly discussions as arguments.

Another challenge to inference confronting our study is **sampling**, or **case selection**—the way in which we select the cases or observations from which we make inferences. The challenge in case selection is choosing cases that will allow us to make valid inferences about the population of interest: in this case, Americans. Frequently, it is not feasible to collect data for the entire population of interest. For example, it is not possible (and certainly not affordable) to measure the attitudes of every single American. In our case, our inference about polarization was generated from observing people we happened to know.

But what threats to inference might this case selection method create for us? One immediate concern we would expect a skeptical social scientist to raise is that people we happen to know are not likely to be representative of the American public in general. Among other differences, given our interest in politics, our acquaintances are probably much more interested in politics than

12 Morris P. Fiorina and Samuel J. Abrams, "Political polarization in the American public," *Annual Review of Political Science* 11 (2008): 563–588.

Americans in general. As a result, we can say that our sample is likely biased. **Bias** is a term we will use in many contexts when discussing challenges to inference; indeed, there are many potential sources of bias when we attempt to make inferences. But regardless of the context, *bias always refers to systematic, or non-random error*. In this case, we likely have constructed a biased sample of Americans, in that our sample will be systematically more interested in politics than the average American. Since previous research teaches us that people who are more interested in politics also tend to have more extreme opinions than those who are less interested, it is likely that the data we collected from our sample of acquaintances will lead us to infer that there is more polarization than we might see if we had surveyed a more representative sample of Americans.

This example may strike you as a relatively straightforward one, but case selection is another crucial step in a research design. How one selects cases is very important and will differ depending on the goal of the study (whether we want to make descriptive or causal inferences) and the type of study (quantitative or qualitative). Thus, we return to the issue of case selection later in this chapter as well as in Chapters 4 and 7.

It should be obvious by now that there are myriad reasons to question the inference we made about political polarization in the United States based on our casual observations of political disagreements among our friends and neighbors. This does not necessarily mean that our conclusion is wrong—it may indeed be true that political polarization has increased. But it does mean that our approach, or our methodology, is too flawed for us to be confident in our conclusion. So how might we improve this research design to reduce the number of challenges we face in making inferences? Let us briefly consider an alternative design that attempts to address some of the limitations of our original approach.

Recall that our first problem occurred because we operationalized our measure in a way that did not match our concept. To address this issue, we can attempt to measure our concept through a survey rather than observation. For example, we could ask people about their opinions on a variety of political issues. Such an operationalization would more directly capture the concept of ideological polarization.

When it comes to measuring this ideological polarization, we still need to be conscious of measurement error. There will certainly be some measurement error when we ask people about their opinions on the issues. Some individuals may not understand the question and inadvertently give the wrong response. It is also possible that the interviewers might record the information incorrectly. These are examples of **random measurement error**. This is called *random* error because it does not systematically bias our results in any one direction. For example, if we were measuring if someone has moderate or extreme views we might randomly overstate one respondent's extremism and understate another's.

By contrast, a greater concern is **systematic measurement error**, which would cause bias in our estimates. For example, systematic measurement error could occur if some respondents thought it was undesirable to express extreme political views and therefore systematically chose more moderate responses than what truly represented their opinions. In this case, our measure would systematically underestimate the amount of polarization that actually existed.

To address concerns about case selection, we could try to construct a sample that is much more representative of the American population. We could do this by taking a random sample of American adults. We discuss this approach in more detail in Chapter 4, but, under a variety of assumptions, a random sample of approximately 1,000 is typically sufficient to be representative of the population, within some margin of sampling error.

In sum, even with an improved research design, it is important to bear in mind that we face many challenges when we seek to make descriptive inferences about the political world. As we discuss in more detail below, the scientific enterprise does not require us to have a flawless research design, but we must endeavor to make inferences in as sound a manner as possible and to be open and transparent about the decisions we make in producing these inferences so that other scholars can evaluate their quality.

Challenges to Causal Inference

Although we face many challenges when we wish to make descriptive inferences about the political world, we must grapple with many more when our goal is to make causal inferences. Imagine that we were able to infer with a reasonable amount of confidence that the American public was more ideologically polarized now than it had been previously. This may lead us to an important research question: What caused Americans to become more polarized? Indeed, this question has been a source of significant debate among political scientists. One hypothesis is that Americans have become more divided as a reaction to polarization among political elites, such as elected officials. We will elaborate more on the formulation of research questions in Chapter 2 and the derivation of theories and hypotheses in Chapter 3. For the moment, it is sufficient to note that this hypothesis was derived by consulting the existing literature on how individuals formulate their opinions; specifically, many scholars have shown that the public tends to take cues from politicians and formulate their opinions accordingly.[13] Thus, it is reasonable to suggest that if politicians begin to express more extreme political views, the public is likely to follow suit.

Of course, this is merely a hypothesis; to be able to make a causal inference we first need to develop a research design that would allow us to test this

13 John Zaller, *The nature and origins of mass opinion* (Cambridge, UK: Cambridge University Press, 1992).

hypothesis. Then, armed with our results, we would be far more confident in making inferences. For example, we could track the polarization of elites and of ordinary Americans over time and see if polarization of the former precedes the latter. Imagine that we used Abramowitz's measure of political polarization among the public—the percentage of people who took a consistently liberal or conservative position on every issue they were asked about. In 2008, Abramowitz and Saunders estimated that about 33 percent of Americans had consistent ideological opinions across all issues.[14] Now suppose that a similar study was conducted several years earlier. At that time elite polarization was shown to be less, and the study found that only 24 percent of Americans held consistent ideological positions on the issues: a difference of 9 percentage points.

Assuming the studies are comparable, it would be tempting to conclude that the causal effect of the independent variable (the polarization of politicians) is 9 percentage points. Of course, this would only be an estimate because we can never observe two alternate states of the world at the same time; that is, we can never know what issue positions Americans would have taken if politicians had not become more polarized over the past few decades. This is the fundamental challenge of causal inference.[15] King, Keohane, and Verba explain just how serious this challenge is: "no matter how perfect the research design, no matter how much data we collect, no matter how perceptive the observers, no matter how diligent the research assistants, and no matter how much experimental control we have, we will never know a causal inference for certain."[16] This statement may seem off-putting to students at first, but our experience is that the most passionate social scientists see this challenge mostly as an exciting one. We can never know for certain whether a causal effect exists, but we *can* design our studies so that we are as certain as possible.

In thinking about how to design our research so that we can make causal inferences with as much confidence as possible, it is always good to keep in mind the **counterfactual**. The counterfactual is the state of the world that you do not observe. In our running example, the counterfactual would be a world where politicians had not become more polarized over the past few decades. While we can never directly observe the counterfactual, we should use the counterfactual to think about how to design a study that would allow us to make reasonable causal inferences. For example, we might think about how we can replicate as closely as possible an environment like our counterfactual, and then use that as a baseline for estimating the effects of our independent variable.

As we noted above, our theory leads us to expect that when politicians become more polarized the public will also become more polarized in response.

14 Abramowitz and Saunders, "Why can't we all just get along?" p. 544.

15 We are paraphrasing Paul W. Holland, "Statistics and causal inference," *Journal of the American Statistical Association* 81 (1986): 945–960, who referred to this as the fundamental problem of causal inference.

16 King, Keohane, and Verba, *Designing social inquiry*, p. 79.

Let's say that to test this theory we compared the polarization of elites and the public in the 1970s with polarization among both groups in the 2000s. Indeed, scholars who have conducted such a comparison have demonstrated convincingly that politicians have become more polarized over that period. And while there is some debate about whether the public has also become more polarized, there are some indications that such polarization has occurred. Thus, there appears to be a relationship between these two variables—when politicians are polarized, so too is the public. Nonetheless, without knowing the counterfactual, a world where no elite polarization occurred, there are good reasons to be skeptical of this finding.

One issue that is often problematic, particularly for observational research, is that of **reverse-causation**. Reverse-causation occurs when the dependent variable causes the independent variable, either instead of or in addition to the expected causal relationship of the independent variable on the dependent variable. Figure 1.1 shows how this might occur with regard to our example; it is possible that rather than the public reacting to increasing polarization among politicians (as shown in panel A), politicians actually become polarized in response to increasing polarization among the public (panel B). This possibility is theoretically plausible, since we know that politicians need support from the public to win elections—so they might very well polarize in response to the public. It is also possible that the relationship is **reciprocal**—that is, that the independent and dependent variables simultaneously impact each other (panel C).[17] This might happen if, for example, elites began to polarize, which then led the public to become more polarized, which then caused politicians to become even more polarized.

The challenge to inference in this case is significant. Imagine that we took a snap shot of polarization among politicians and the public in the 1970s and then we took another snap shot of polarization among both groups in the 2000s. The problem is that our expected causal relationship (polarization among politicians causes polarization among the public) would yield identical results as the reverse relationship (polarization among the public causes polarization among politicians); regardless of whether politicians caused the public to polarize or vice versa, both would be more polarized in the 2000s than they were in the 1970s. Thus, we cannot confidently say that elite polarization has caused polarization among the mass public if we cannot rule out the possibility that the relationship exists in the opposite direction.

Another challenge to causal inference that we must often confront is called **spuriousness**, a dynamic that occurs when some other factor causes changes in both the dependent and independent variables simultaneously. Indeed, even if two variables move together, this does not always mean that they are directly related to each other. For example, since the 1970s, the mass media

17 A reciprocal relationship is also sometimes referred to as a recursive relationship.

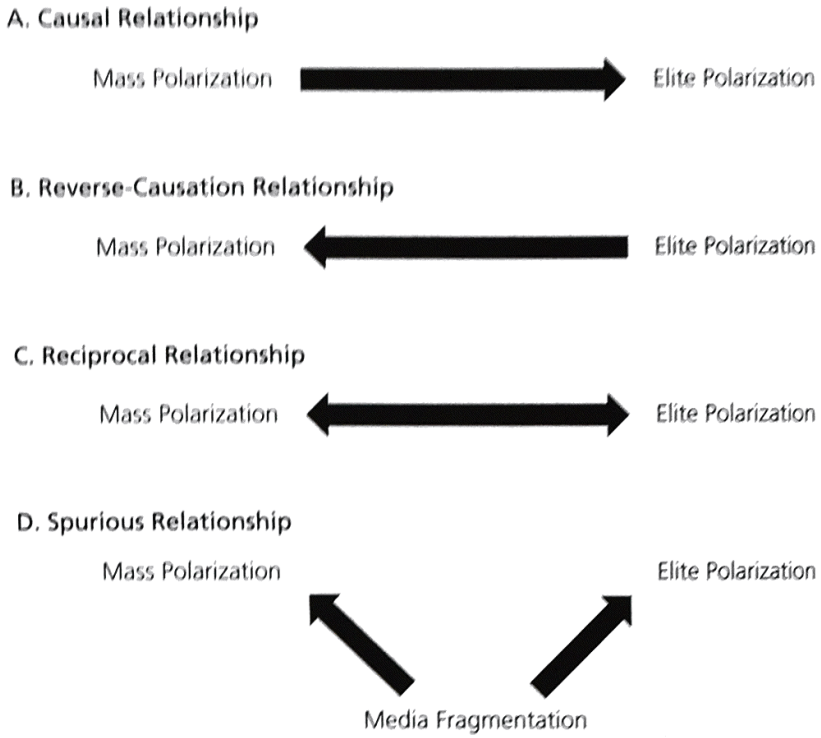

FIGURE 1.1 Different Explanations for Increasing Polarization among Elites and the Public

has become more highly fragmented, leading news outlets to cater to specific ideological outlooks (for example, Fox News for conservatives and MSNBC for liberals). This change in the nature of the modern news media may produce mass polarization by limiting the extent to which ideologues hear opposing viewpoints; at the same time; it may also produce incentives for politicians to take more extreme political views to appeal to these ideological news outlets (shown in Figure 1.1, panel D). In this case, the relationship we thought we detected between elite and mass polarization would actually be spurious, with both types of polarization being caused by media fragmentation instead.

Threats to inference such as reverse-causation and spuriousness must first be identified before they can be systematically addressed and (hopefully) ruled out as alternative explanations for the relationships we observe. Once again, theory is a crucial step in identifying possible reverse-causation and spurious relationships. Without a strong sense of what political scientists have learned about how the world may operate, it is less likely to occur to us that the relationship between elite polarization and mass polarization might be spurious. Once these alternative explanations have been identified, it may be possible for us to design our research in a way that systematically addresses these alternative explanations. For example, to address the possibility of reverse-causation, we may decide to measure polarization every year since 1970 rather than just at two points in time. With such an approach, we would be better able to

determine whether the public's attitudes started to become more polarized before politicians or vice versa. We could also attempt to control for variables that may be causing spurious relationships; for example, perhaps we could track polarization among people who do not have access to cable television, thereby reducing the possible spurious relationship between media fragmentation and polarization.

Our approach to addressing the challenges of potential spurious variables or reverse-causation generally depends on the approach we are taking in our study. We discuss these approaches more specifically in Chapters 4–7. But one of the most important steps is simply being aware of these challenges in the first place.

THE IMPORTANCE OF THEORY

The ability to make robust causal inferences depends on a scholar's development of strong theoretical explanations that can support the causal processes she is asserting. Human beings are very good at recognizing patterns in the world, and, if we look at enough data, we are likely to see relationships simply due to happenstance. For many years, journalists noted an interesting pattern between World Series victories and presidential election outcomes. Specifically, from 1952 to 1976, whenever a team from the American League won the World Series in October, the Republican candidate won the presidential election that November. Whenever the National League won the World Series, the Democrat was the winner. Thus, there appeared to be a strong association between World Series outcomes and presidential election outcomes. Yet, one could hardly make the causal inference that the World Series determined the election results, largely because there is no good theory, or explanation, for why this would be so. Indeed, since 1976 there has been no strong relationship between the results of the World Series and the presidential election; the break in the pattern is hardly surprising given that we had no theoretical justification for expecting the two things to be related in the first place.

Post-Hoc Theory

The example of baseball results and election outcomes is an extreme one, but the issue of unidentified or under-identified causal theories is real for many research projects. One well-debated example comes from international relations in the form of the "Democratic Peace Theory."[18] The Democratic Peace Theory is not really a single theory, but rather a collection of post-hoc, or

18 For a review of the research in this area see James Lee Ray, "Does democracy cause peace?" *Annual Review of Political Science* 1 (1998): 27–46.

after-the-fact, theories developed to explain the persistent empirical finding that nations with democratic governments rarely engage in military conflict against each other. The two most prominent theories given for this empirical regularity is that (1) democracies share common democratic norms, which prevent them from coming into conflict with each other, and that (2) democratic institutions constrain leaders from engaging in such conflicts. However, these explanations have come under some criticism in the past several years. For example, Sebastian Rosato argues that "the causal logics that underpin the democratic peace theory cannot explain why democracies remain at peace with one another because the mechanisms that make up these logics do not operate as stipulated by the theory's proponents."[19] With regard to the first theory, Rosato points out that democracies frequently violate democratic norms when they carry out their foreign policy. One of the most consistent examples of this pattern is the extent to which European democracies were willing to engage in wars to maintain their autocratic rule over their colonies.

Rosato also calls into question the second theory for why democracies do not go to war—the fact that democratic leaders may be held more accountable for engaging in such conflicts. Rosato examines how frequently democratic leaders are removed from office following a costly war compared with the frequency with which dictators lose their positions under similar circumstances. Rosato shows that dictators actually lose their office at least as frequently as democratic leaders following a "costly war." Thus, the notion that democratic leaders will feel more constrained when deciding whether to engage in a conflict appears to be very questionable. As a result, Rosato finds that the democratic peace theory is not particularly convincing because the causal mechanisms specified by the theory do not appear to operate in the manner expected. Rosato's findings illustrate the importance of having a strong theory, or explanation, for why an independent variable causes a dependent variable.[20]

Deductive vs. Inductive Approaches

We do not mean to suggest that a relationship cannot be causal if scholars did not theorize a causal process in advance. As we will stress throughout this book, the research process is not an orderly one; often we must revisit our research

19 Sebastian Rosato, "The flawed logic of democratic peace theory," *American Political Science Review* 97 (2003): 585–602, p. 599.

20 Other research has suggested that the relationship between democracies and peace may be spurious. Democracies tend to join alliances and they also tend to have stronger economic ties. Nations that have alliances or are reliant on each other economically rarely fight each other. See Henry S. Farber and Joanne Gowa, "Polities and peace," *International Security* 20 (1995): 123–146; Henry S. Farber and Joanne Gowa, "Common interests or common polities? Reinterpreting the democratic peace," *Journal of Politics* 59 (1997): 393–417; Erik Gartzke, "The capitalist peace," *American Journal of Political Science* 51 (2007): 166–191.

question and our theories and hypotheses once we analyze the data. Indeed, it is often the case that we discover relationships through the analysis of collected data, and then we must rethink why those relationships may exist. It may be that there is a very good reason to expect a causal relationship but that we had simply not recognized that reasoning in advance of conducting our research.

In fact, one of the most robust debates in the social sciences has focused on this issue. **Deductive research** follows the approach outlined above: scholars begin with a research question, develop theory to provide an answer to that question, derive hypotheses from that theory to be tested, and develop and implement a research design to test these hypotheses. **Inductive research**, however, reverses this order. Rather than presume to know how the world works and test hypotheses, inductive research starts by observing the natural world and then derives hypotheses and theories from these observations.

Inferences based on an inductive approach are more likely to suffer from the challenges to causal inference discussed above. Continuing the example, an observed relationship between elite polarization and public polarization could be spurious if media fragmentation isn't taken into account. A deductive study, however, could be designed in such a way so as to control for variation in the type of media. Here again lies the benefit of building on pre-existing scholarship. Rather than just observe a topic for the first time, deductive research approaches can build on the findings of previous scholarship. Nonetheless, inductive approaches are particularly helpful for exploratory research on questions that have not been previously addressed or have not been well addressed by existing scholarship. Then the hypotheses generated through such an inductive approach can be more systematically tested through a well-designed deductive study.

SUMMARIZING COMMON THREATS TO INFERENCE

So far in this chapter, we have elaborated on many challenges to descriptive and causal inferences. We will continue to return to these and other inferential challenges throughout the text, as these challenges are what motivate our lessons regarding research design. However, we summarize the main threats to inference here as well as provide a road map to where solutions to these challenges are addressed in this book:

- Operationalization of concepts
  - Problem: If we measure our concepts in a way that is not consistent with how we defined those concepts in the first place, then the inferences we draw from the variables may not truly apply to our theoretical concepts.
  - Solution: Make careful connections between our theory and how we measure our concepts (Chapter 3).

- Case selection
  - Problem: If the cases we choose to study are not representative of the world we wish to make inferences about, then our descriptive inferences may not be valid.
  - Solution: Select cases systematically to ensure inferences are as valid as possible (Chapters 4 and 7).
- Measurement error
  - Problem: If there is a lot of error in how we measure our concepts, then we can be less certain about our descriptive and causal inferences. In addition, if our measurement error is systematic, then we run the risk of making biased inferences, inferences that are systematically wrong.
  - Solution: Measure concepts in a way that minimizes systematic and random measurement error (Chapter 4).
- Reverse-causation
  - Problem: In many cases, it is possible that the dependent variable may actually be causing the independent variable, rather than the other way around.
  - Solution: Attempt to rule out the possibility of reverse-causation, possibly by leveraging information about how the variables change over time (Chapter 6) or by using an experimental design (Chapter 5).
- Spurious relationships
  - Problem: The relationship we detect between an independent variable and the dependent variable may be caused by some third variable that causes changes in both variables at the same time.
  - Solution: In an observational design, control for the variable that may be causing the spurious results (Chapter 6). Alternatively, use an experimental design to reduce the likelihood of spuriousness (Chapter 5).

EVALUATING INFERENCES

We noted above that people make inferences about the world every day, but it is critical to note that not all inferences are equally valid—that is, they are not all equally likely to be correct. This is where science comes in. Science is the process of making inferences in a systematic way. This process requires a disciplined and self-conscious approach. Making inferences scientifically entails

a focused research question, theory that builds on previous research, testable hypotheses, and a detailed and systematic plan for collecting and analyzing data that can be used to test these hypotheses. Good social science goes a step further and ensures that the methods followed are made public so that others can judge the quality of those inferences.

The last point is perhaps as important as any: using a scientific approach to make inferences does not ensure that those inferences will be correct, but the fact that the procedures are public does allow other scholars to evaluate how likely it is that those inferences are valid. To ensure that the procedures taken in any study are public, every study should include a section that details as precisely as possible how the data were collected, how the variables were measured, and how the data were analyzed. In fact, any research paper should include enough information so that the reader could reproduce the study based solely on the information provided in the paper.

Box 1.2: Transparency, Replication, and Ethical Social Science

Social scientists are not often the focus of a flurry of news stories (even if their findings occasionally are), but in 2011 a media storm erupted around one prominent psychologist. The psychologist in question was a giant in his field—his research had been frequently published in the top academic journals and often received coverage in prominent news outlets such as the *New York Times*. Unfortunately, in 2011, it was discovered that most of these research studies were fraudulent. According to reports, the psychologist had rarely actually carried out the experiments that his articles were based on; rather, he simply made up data that would support his hypotheses and then published results from that manufactured data.[a]

When the professor's behavior was finally uncovered, there was outrage from social scientists across all disciplines. After all, to build knowledge in any field requires that we be able to trust that published studies are based on a scientific approach that is both valid and replicable. While nearly all social scientists are interested first and foremost in making valid inferences, there are a rare few who appear tempted to fabricate results in the pursuit of bolstering their own stature within the discipline. This is just one of many reasons why the social sciences have increasingly emphasized the importance of the replicability of published research. In political science, many of the top journals now require that when an article is published the authors of the article must make the data available to the scholarly community along with instructions on how the researchers analyzed the data. Such transparency and openness will not only allow scholars to be more confident in the results presented in a particular study, but it may also dissuade scholars from engaging in improper behavior.

a. See, for example, Ewen Callaway, "Report finds massive fraud at Dutch universities," *Nature* 479, November 3, 2011.

KEY TERMS

bias 22
case selection 21
causal inference 18
counterfactual 24
deductive research 29
dependent variable 16
descriptive inference 18
experimental studies 17
hypotheses 16
independent variable 16
inductive research 29
inference 18
measurement error 21
observational study 17
operationalization 16
random measurement error 22
reciprocal 25
research question 15
reverse-causation 25
sampling 21
spuriousness 25
systematic measurement error 23
theory 16
variable 16