1 The Scientific Study of Politics

OVERVIEW

Most political science students are interested in the substance of politics and not in its methodology. We begin with a discussion of the goals of this book and why a scientific approach to the study of politics is more interesting and desirable than a "just-the-facts" approach. In this chapter we provide an overview of what it means to study politics scientifically. We begin with an introduction to how we move from causal theories to scientific knowledge, and a key part of this process is thinking about the world in terms of *models* in which the concepts of interest become variables that are causally linked together by theories. We then introduce the goals and standards of political science research that will be our rules of the road to keep in mind throughout this book. The chapter concludes with a brief overview of the structure of this book.

Doubt is the beginning, not the end, of wisdom.
– Chinese proverb

1.1 POLITICAL *SCIENCE*?

"Which party do you support?" "When are you going to run for office?" These are questions that students often hear after announcing that they are taking courses in political science. Although many political scientists are avid partisans, and some political scientists have even run for elected offices or have advised elected officials, for the most part this is not the focus of modern political science. Instead, political science is about the scientific study of political phenomena. Perhaps like you, a great many of today's political scientists were attracted to this discipline as undergraduates because of intense interests in a particular issue or candidate. Although we are often drawn into political science based on political passions, the most respected political science research today is conducted in a fashion that makes it impossible to tell the personal political views of the writer.

Many people taking their first political science research course are surprised to find out how much science and, in particular, how much math are involved. We would like to encourage the students who find themselves in this position to hang in there with us – even if your answer to this encouragement is "but I'm only taking this class because they require it to graduate, and I'll never use any of this stuff again." Even if you never run a regression model after you graduate, having made your way through these materials should help you in a number of important ways. We have this written this book with the following three goals in mind:

- *To help you consume academic political science research in your other courses.* One of the signs that a field of research is becoming scientific is the development of a common technical language. We aim to make the common technical language of political science accessible to you.
- *To help you become a better consumer of information.* In political science and many other areas of scientific and popular communication, claims about causal relationships are frequently made. We want you to be better able to evaluate such claims critically.
- *To start you on the road to becoming a producer of scientific research on politics.* This is obviously the most ambitious of our goals. In our teaching we often have found that once skeptical students get comfortable with the basic tools of political science, their skepticism turns into curiosity and enthusiasm.

To see the value of this approach, consider an alternative way of learning about politics, one in which political science courses would focus on "just the facts" of politics. Under this alternative way, for example, a course offered in 1995 on the politics of the European Union (EU) would have taught students that there were 15 member nations who participated in governing the EU through a particular set of institutional arrangements that had a particular set of rules. An obvious problem with this alternative way is that courses in which lists of facts are the only material would probably be pretty boring. An even bigger problem, though, is that the political world is constantly changing. In 2008 the EU is made up of 27 member nations and has some new governing institutions and rules that are different from what they were in 1995. Students who took a facts-only course on the EU back in 1995 would find themselves lost in trying to understand the EU of 2008. By contrast, a theoretical approach to politics helps us to better understand why changes have come about and their likely impact on EU politics.

In this chapter we provide an overview of what it means to study politics scientifically. We begin this discussion with an introduction to how we move from causal theories to scientific knowledge. A key part of this process is thinking about the world in terms of *models* in which the concepts of interest become **variables**[1] that are causally linked together by theories. We then introduce the goals and standards of political science research that will be our rules of the road to keep in mind throughout this book. We conclude this chapter with a brief overview of the structure of this book.

1.2 APPROACHING POLITICS SCIENTIFICALLY: THE SEARCH FOR CAUSAL EXPLANATIONS

I've said, I don't know whether it's addictive. I'm not a doctor. I'm not a scientist.

– Bob Dole, in a conversation with Katie Couric about tobacco during the 1996 U.S. presidential campaign

The question of "how do we know what we know" is, at its heart, a philosophical question. Scientists are lumped into different disciplines that develop standards for evaluating evidence. A core part of being a scientist and taking a scientific approach to studying the phenomena that interest you is always being willing to consider new evidence and, on the basis of that new evidence, change what you thought you *knew* to be true. This willingness to always consider new *evidence is counterbalanced* by a stern approach to the evaluation of new evidence that permeates the scientific approach. This is certainly true of the way that political scientists approach politics.

So what do political scientists do and what makes them scientists? A basic answer to this question is that, like other scientists, political scientists develop and test theories. A **theory** is a tentative conjecture about the causes of some phenomenon of interest. Once a theory has been developed, we can restate it into one or more testable hypotheses. A **hypothesis** is a theory-based statement about a relationship that we expect to observe. For every hypothesis there is a corresponding **null hypothesis**. A null hypothesis is also a theory-based statement but it is about what we would expect to observe if our theory was incorrect. **Hypothesis testing** is a process in which scientists evaluate systematically collected evidence to make a judgement of

[1] When we introduce an important new term in this book, that term appears in boldface type. We discuss variables at great length later in this and other chapters. For now, a good working definition is that a variable is something that varies. An example of a variable is voter turnout; researchers usually measure it as the percentage of voting-eligible persons in a geographically defined area who cast a vote in a particular election.

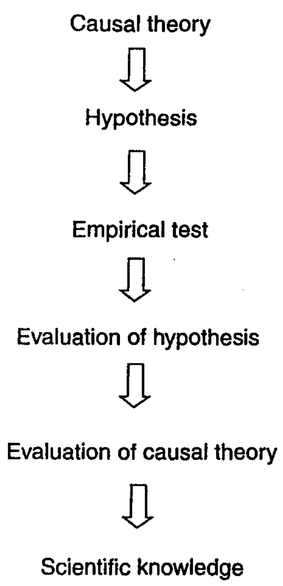

Figure 1.1. The road to scientific knowledge.

whether the evidence favors their hypothesis or favors the corresponding null hypothesis. If a hypothesis survives a series of rigorous tests, scientists start to gain confidence in that hypothesis rather than in the null hypothesis, and thus they also gain confidence in the theory from which they generated their hypothesis.

Figure 1.1 presents a stylized schematic view of the path from theories to hypotheses to scientific knowledge.[2] At the top of the figure, we begin with a **causal theory** to explain our phenomenon of interest. We then derive one or more hypotheses about what our theory leads us to expect when we measure our concepts of interest (which we call **variables** – as subsequently discussed) in the real world. In the third step, we conduct **empirical tests** of our hypotheses.[3] From what we find, we evaluate our hypotheses relative to corresponding null hypotheses. Next, from the results of our hypothesis tests, we evaluate our causal theory. In light of our evaluation of our theory, we then think about how, if at all, we should revise what we consider to be scientific knowledge concerning our phenomenon of interest.

A core part of the scientific process is skepticism. On hearing of a new theory, other scientists will challenge this theory and devise further tests. Although this process can occasionally become quite combative, it is a necessary component in the development of scientific knowledge. Indeed, a core component of scientific knowledge is that, as confident as we are in a particular theory, we remain open to the possibility that there is still a test out there that will provide evidence that makes us lose confidence in that theory.

It is important to underscore here the nature of the testing that scientists carry out. One way of explaining this is to say that scientists are *not* like lawyers in the way that they approach evidence. Lawyers work for a particular client, advocate a particular point of view (like "guilt" or "innocence"), and then accumulate evidence with a goal of proving their case to a judge or jury. This goal of *proving* a desired result determines

[2] In practice, the development of scientific knowledge is frequently much messier than this step-by-step diagram. We show more of the complexity of this approach in later chapters.
[3] By "empirical" we simply mean "based on observations of the real world."

their approach to evidence. When faced with evidence that conflicts with their case, lawyers attempt to ignore or discredit such evidence. When faced with evidence that supports their case, lawyers try to emphasize the applicability of the supportive evidence. In many ways, the scientific and legal approaches to evidence couldn't be further apart. Scientific confidence in a theory is achieved only after hypotheses derived from that theory have run a gantlet of tough tests. At the beginning of a trial, lawyers develop a strategy to *prove* their case. In contrast, at the beginning of a research project, scientists will think long and hard about the most rigorous tests that they can conduct. A scientist's theory is never *proven* because scientists are always willing to consider new evidence.

The process of hypothesis testing reflects how hard scientists are on their own theories. As scientists evaluate systematically collected evidence to make a judgment of whether the evidence favors their hypothesis or favors the corresponding null hypothesis, they *always* favor the null hypothesis. Statistical techniques allow scientists to make probability-based statements about the empirical evidence that they have collected. You might think that, if the evidence was 50–50 between their hypothesis and the corresponding null hypothesis, the scientists would tend to give the nod to the hypothesis (from their theory) over the null hypothesis. In practice, though, this is not the case. Even when the hypothesis has an 80–20 edge over the null hypothesis, most scientists will still favor the null hypothesis. Why? Because scientists are very worried about the possibility of falsely rejecting the null hypothesis and therefore making claims that others ultimately will show to be wrong.

Once a theory has become established as a part of scientific knowledge in a field of study, researchers can build upon the foundation that this theory provides. Thomas Kuhn wrote about these processes in his famous book *The Structure of Scientific Revolutions*. According to Kuhn, scientific fields go through cycles of accumulating knowledge based on a set of shared assumptions and commonly accepted theories about the way that the world works. Together, these shared assumptions and accepted theories form what we call a **paradigm**. Once researchers in a scientific field have widely accepted a paradigm, they can pursue increasingly technical questions that make sense only because of the work that has come beforehand. This state of research under an accepted paradigm is referred to as **normal science**. When a major problem is found with the accepted theories and assumptions of a scientific field, that field will go through a revolutionary period during which new theories and assumptions replace the old paradigm to establish a new paradigm. One of the more famous of these scientific revolutions occurred during the 16th century when the field of astronomy was forced to abandon its assumption that the Earth was the center of the known universe. This was an assumption that had informed theories about planetary movement for thousands of years. In the book *On Revolutions of the Heavenly Bodies*, Nicolai Copernicus presented his theory that the Sun was the center of the known universe. Although this radical theory met many challenges, an increasing body of evidence convinced astronomers that Coperinicus had it right. In the aftermath of this paradigm shift, researchers developed new assumptions and theories that established a new paradigm, and the affected fields of study entered into new periods of normal scientific research.

It may seem hard to imagine that the field of political science has gone through anything that can compare with the experiences of astronomers in the 16th century. Indeed, Kuhn and other scholars who study the evolution of scientific fields of research have a lively and ongoing debate about where the social sciences, like political science, are in terms of their development. The more skeptical participants in this debate argue that political science is not sufficiently mature to have a paradigm, much less a paradigm shift. If we put aside this somewhat esoteric debate about paradigms and paradigm shifts, we can see an important example of the evolution of scientific knowledge about politics from the study of public opinion in the United States.

In the 1940s the study of public opinion through mass surveys was in its infancy. Prior to that time, political scientists and sociologists assumed that U.S. voters were heavily influenced by presidential campaigns – and, in particular, by campaign advertising – as they made up their minds about the candidates. To better understand how these processes worked, a team of researchers from Columbia University set up an in-depth study of public opinion in Erie County, Ohio, during the 1944 presidential election. Their study involved interviewing the same individuals at multiple time periods across the course of the campaign. Much to the researchers' surprise, they found that voters were remarkably consistent from interview to interview in terms of their vote intentions. Instead of being influenced by particular events of the campaign, most of the voters surveyed had made up their minds about how they would cast their ballots long before the campaigning had even begun. The resulting book by Paul Lazarsfeld, Bernard Berelson, and Hazel Gaudet, titled *The People's Choice*, changed the way that scholars thought about public opinion and political behavior in the United States. If political campaigns were not central to vote choice, scholars were forced to ask themselves what *was* critical to determining how people voted.

At first other scholars were skeptical of the findings of the 1944 Erie County study, but as the revised theories of politics of Lazarsfeld et al. were evaluated in other studies, the field of public opinion underwent a change

that looks very much like what Thomas Kuhn calls a "paradigm shift." In the aftermath of this finding, new theories were developed to attempt to explain the origins of voters' long-lasting attachments to political parties in the United States. An example of an influential study that was carried out under this shifted paradigm is Richard Niemi and Kent Jenning's seminal book from 1974, *The Political Character of Adolescence: The Influence of Families and Schools*. As the title indicates, Niemi and Jennings studied the attachments of schoolchildren to political parties. Under the pre–Erie County paradigm of public opinion, this study would not have made much sense. But once researchers had found that voter's partisan attachments were quite stable over time, studying them at the early ages at which they form became a reasonable scientific enterprise. You can see evidence of this paradigm at work in current studies of party identification and debates about its stability.

1.3 THINKING ABOUT THE WORLD IN TERMS OF VARIABLES AND CAUSAL EXPLANATIONS

So how do political scientists develop theories about politics? A key element of this is that they order their thoughts about the political world in terms of concepts that scientists call *variables* and causal relationships between variables. This type of mental exercise is just a more rigorous way of expressing ideas about politics that we hear on a daily basis. You should think of each variable in terms of its *label* and its *values*. The ***variable label*** is a description of what the variable is, and the **variable values** are the denominations in which the variable occurs. So, if we're talking about the variable that reflects an individual's age, we could simply label this variable "Age" and some of the denominations in which this variable occurs would be years, days, or even hours.

It is easier to understand the process of turning concepts into variables by using an example of an entire theory. For instance, if we're thinking about U.S. presidential elections, a commonly expressed idea is that the incumbent president will fare better when the economy is relatively healthy. If we restate this in terms of a political science theory, the state of the economy becomes the **independent variable**, and the outcome of presidential elections becomes the **dependent variable**. One way of keeping the lingo of theories straight is to remember that the value of the "dependent" variable "depends" on the value of the "independent" variable. Recall that a theory is a tentative conjecture about the causes of some phenomenon of interest. In other words, a theory is a conjecture that the independent variable is causally related to the dependent variable; according to our theory, change in the value of the independent variable *causes* change in the value of the dependent variable.

This is a good opportunity to pause and try to come up with your own causal statement in terms of an independent and dependent variable; try filling in the following blanks with some political variables:

________________ **causes** ________________

Sometimes it's easier to phrase causal propositions more specifically in terms of the values of the variables that you have in mind. For instance,

higher ________________ **causes lower** ________________

or

higher ________________ **causes higher** ________________

Once you learn to think about the world in terms of variables you will be able to produce an almost endless slew of causal theories. In Chapter 4 we will discuss at length how we design research to evaluate the causal claims in theories, but one way to initially evaluate a particular theory is to think about the **causal explanation** behind it. The causal explanation behind a theory is the answer to the question, "why do you think that this independent variable is causally related to this dependent variable?" If the answer is reasonable, then the theory has possibilities. In addition, if the answer is original and thought provoking, then you may really be onto something. Let's return now to our working example in which the state of the economy is the independent variable and the outcome of presidential elections is our dependent variable. The causal explanation for this theory is that we believe that the state of the economy is *causally related* to the outcome of presidential elections *because* voters hold the president responsible for management of the national economy. As a result, when the economy has been performing well, more voters will vote for the incumbent. When the economy is performing poorly, fewer voters will support the incumbent candidate. If we put this in terms of the preceding fill-in-the-blank exercise, we could write

economic performance causes presidential election outcomes,

or, more specifically, we could write

higher economic performance causes higher incumbent vote.

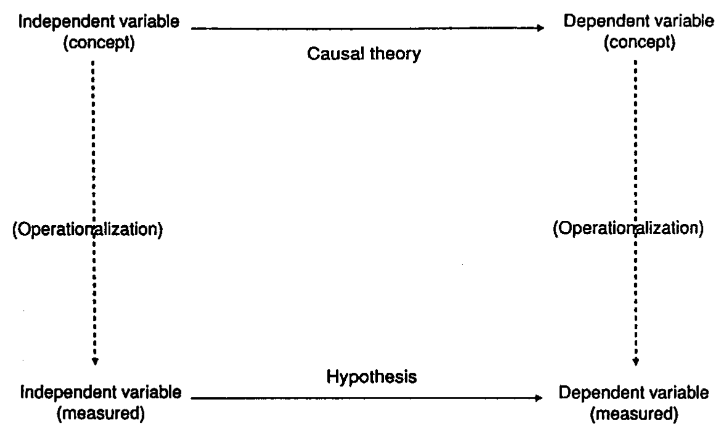

Figure 1.2. From theory to hypothesis.

For now we'll refer to this theory, which has been widely advanced and tested by political scientists, as "the theory of economic voting."

To test the theory of economic voting in U.S. presidential elections, we need to derive from it one or more testable hypotheses. Figure 1.2 provides a schematic diagram of the relationship between a theory and one of its hypotheses. At the top of this diagram are the components of the causal theory. As we move from the top part of this diagram (Causal theory) to the bottom part (Hypothesis), we are moving from a general statement about how we think the world works to a more specific statement about a relationship that we expect to find when we go out in the real world and measure (or **operationalize**) our variables.[4]

At the theory level at the top of Figure 1.2, our variables do not need to be explicitly defined. With the economic voting example, the independent variable, "Economic Performance," can be thought of as a concept that ranges from very strong to very poor. The dependent variable, "Incumbent Vote," can be thought of as a concept that ranges from very high to very low. Our causal theory is that a stronger economic performance causes the incumbent vote to be higher.

Because there are many ways in which we can measure each of our two variables, there are many different hypotheses that we can test to find out how well our theory holds up to real-world data. We can measure economic performance in a variety of ways. These measures include inflation,

[4] Throughout this book we will use the terms "measure" and "operationalize" interchangeably. It is fairly common practice in the current political science literature to use the term "operationalize."

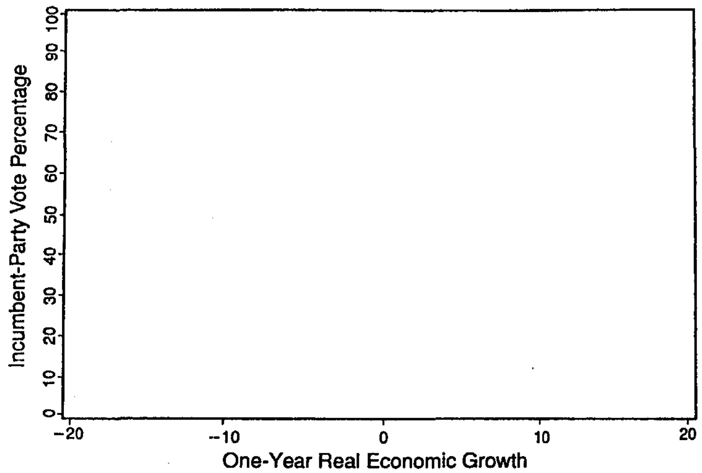

Figure 1.3. What would you expect to see based on the theory of economic voting?

unemployment, real economic growth, and many others. "Incumbent Vote" may seem pretty straightforward to measure, but here there are also a number of choices that we need to make. For instance, what do we do in the cases in which the incumbent president is not running again? Or what about elections in which a third-party candidate runs? Measurement (or operationalization) of concepts is an important part of the scientific process. We will discuss this in greater detail in Chapter 5, which is devoted entirely to variable measurement. For now, we imagine that we are operationalizing economic performance with real economic growth, as defined by official U.S. government measures of the one-year rate of inflation-adjusted economic growth at the time of the election. We operationalize our dependent variable as the percentage of the popular vote, as reported in official election results, for the party that controlled the presidency at the time of the election.

Figure 1.3 shows the axes of the graph that we could produce if we collected the measures of these two variables. We could place each U.S. presidential election on the graph in Figure 1.3 by identifying the point that corresponds to the value of both "One-Year Real Economic Growth" (the horizontal, or x, axis) and "Incumbent-Party Vote Percentage" (the vertical, or y, axis). For instance, if these values were (respectively) 0 and 50, the position for that election year would be exactly in the center of the graph. Based on our theory, what would you expect to see if we collected these measures for all elections? Remember that our theory is that a stronger *economic performance* causes the *incumbent vote* to be higher. And we can restate

this theory in reverse such that a weaker *economic performance* causes the *incumbent vote* to be lower. So, what would this lead us to expect to see if we plotted real-world data onto Figure 1.3? To get this answer right, let's make sure that we know our way around this graph. If we move from left to right on the horizontal axis, which is labeled "One-Year Real Economic Growth," what is going on in real-world terms? We can see that, at the far left end of the horizontal axis, the value is −20. This would mean that the U.S. economy had shrunk by 20% over the past year, which would represent a very poor performance (to say the least). As we move to the right on this axis, each point represents a better economic performance up to the point where we see a value of +20, indicating that the real economy has grown by 20% over the past year. The vertical axis depicts values of "Incumbent-Party Vote Percentage." Moving upward on this axis represents an increasing share of the popular vote for the incumbent party, whereas moving downward represents a decreasing share of the popular vote.

Now think about these two axes together in terms of what we would expect to see based on the theory of economic voting. In thinking through these matters, we should always start with our independent variable. This is because our theory states that the value of the independent variable exerts a causal influence on the value of the dependent variable. So, if we start with a very low value of *economic performance* – let's say −15 on the horizontal axis – what does our theory lead us to expect in terms of values for the *incumbent vote*, the dependent variable? We would also expect the value of the dependent variable to be very low. This case would then be expected to be in the lower-left-hand corner of Figure 1.3. Now imagine a case in which economic performance was quite strong at +15. Under these circumstances, our theory would lead us to expect that the incumbent-vote percentage would also be quite high. Such a case would be in the upper-right-hand corner of our graph. Figure 1.4 shows two such hypothetical points plotted on the same graph as Figure 1.3. If we draw a line between these two points, this line would slope upward from the lower left to the upper right. We describe such a line as having a positive slope. We can therefore hypothesize that the relationship between the variable labeled "One-Year Real Economic Growth" and the variable labeled "Incumbent-Party Vote Percentage" will be a positive relationship. A positive relationship is one for which higher values of the independent variable coincide with higher values of the dependent variable.

Let's consider a different operationalization of our independent variable. Instead of economic growth, let's use "Unemployment Percentage" as our operationalization of economic performance. We haven't changed our theory, but we need to rethink our hypothesis with this new measurement or operationalization. The best way to do so is to draw a picture like

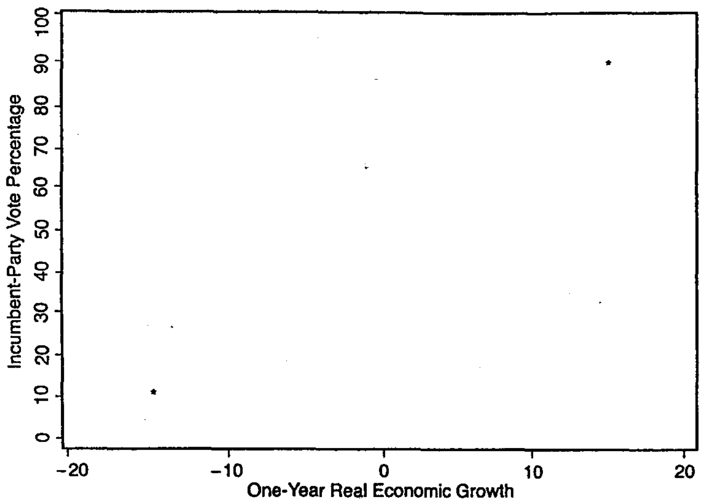

Figure 1.4. What would you expect to see based on the theory of economic voting? Two hypothetical cases.

Figure 1.3 but with the changed independent variable on the horizontal axis. This is what we have in Figure 1.5. As we move from left to right on the horizontal axis in Figure 1.5, the percentage of the members of the workforce who are unemployed goes up. What does this mean in terms

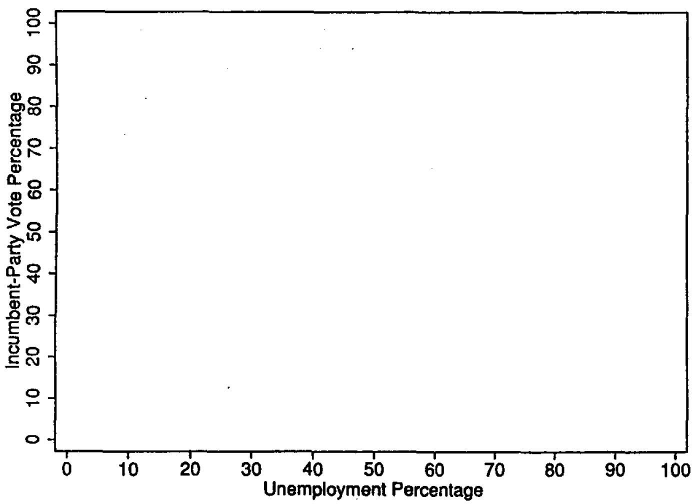

Figure 1.5. What would you expect to see based on the theory of economic voting?

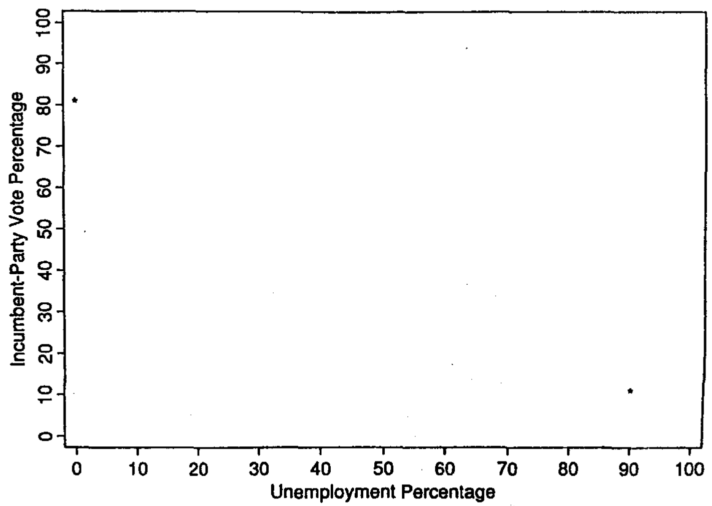

Figure 1.6. What would you expect to see based on the theory of economic voting? Two hypothetical cases.

of economic performance? Rising unemployment is generally considered a poorer economic performance whereas decreasing unemployment is considered a better economic performance. Based on our theory, what should we expect to see in terms of incumbent vote percentage when unemployment is high? What about when unemployment is low?

Figure 1.6 shows two such hypothetical points plotted on our graph of unemployment and incumbent vote from Figure 1.5. The point in the upper-left-hand corner represents our expected vote percentage when unemployment equals zero. Under these circumstances, our theory of economic voting leads us to expect that the incumbent party will do very well. The point in the lower-right-hand corner represents our expected vote percentage when unemployment is very high. Under these circumstances our theory of economic voting leads us to expect that the incumbent party will do very poorly. If we draw a line between these two points, this line would slope downward from the upper-left to the lower-right. We describe such a line as having a negative slope. We can therefore hypothesize that the relationship between the variable labeled "Unemployment Percentage" and the variable labeled "Incumbent-Party Vote Percentage" will be a negative relationship. A negative relationship is one for which higher values of the independent variable coincide with lower values of the dependent variable.

In this example we have seen that the same theory can lead to a hypothesis of a positive or a negative relationship. The operationalization of the independent and the dependent variables determines the direction of the hypothesized relationship. It is often very helpful to draw a picture like Figure 1.3 or 1.5 to translate our theories into hypotheses. Once we have such a figure with the axes properly labeled, we can determine what our expected value of our dependent variable should be if we observe both a high and a low value of the independent variable. And once we have placed the two resulting points on our figure, we can tell whether our hypothesized relationship is positive or negative.

Once we have figured out our hypothesized relationship, we can collect data from real-world cases and see how well these data reflect our expectations of a positive or negative relationship. This is a very important step that we can carry out fairly easily in the case of the theory of economic voting. Once we collect all of the data on economic performance and election outcomes, we will, however, still be a long way from confirming the theory that economic performance *causes* presidential election outcomes. Even if a graph like Figure 1.3 produces compelling visual evidence, we will need to see more rigorous evidence than that. Chapters 8–12 focus on the evaluation of hypotheses by use of statistics. The basic logic of statistical hypothesis testing is that we assess the probability that the relationship we find could be due to random chance. The stronger the evidence that such a relationship *could not* be due to random chance, the more confident we would be in our hypothesis. The stronger the evidence that such a relationship *could* be due to random chance, the more confident we would be in the corresponding null hypothesis. This in turn reflects on our theory.

We also, at this point, need to be cautious about claiming that we have "confirmed" our theory, because social scientific phenomena (such as elections) are usually complex and cannot be explained completely with a single independent variable. Take a minute or two to think about what other variables, aside from economic performance, you believe might be causally related to U.S. presidential election outcomes. If you can come up with at least one, you are on your way to thinking like a political scientist. Because there are usually other variables that matter, we can continue to think about our theories two variables at a time, but we need to qualify our expectations to account for other variables. We will spend Chapters 3 and 4 expanding on these important issues.

1.4 MODELS OF POLITICS

When we think about the phenomena that we want to better understand as dependent variables and develop theories about the independent variables that causally influence them, we are constructing **theoretical models**.

Political scientist James Rogers provides an excellent analogy between models and maps to explain how these abstractions from reality are useful to us as we try to understand the political world:

> ... the very unrealism of a model, if properly constructed, is what makes it useful. The models developed below are intended to serve much the same function as a street map of a city. If one compares a map of a city to the real topography of that city, it is certain that what is represented in the map is a highly unrealistic portrayal of what the city actually looks like. The map utterly distorts what is *really* there and leaves out numerous details about what a particular area looks like. But it is precisely *because* the map distorts reality – because it abstracts away from a host of details about what is really there – that it is a useful tool. A map that attempted to portray the full details of a particular area would be too cluttered to be useful in finding a particular location or would be too large to be conveniently stored. (2006, p. 276, emphasis in original)

The essential point is that models *are* simplifications. Whether or not they are useful to us depends on what we are trying to accomplish with the particular model. One of the remarkable aspects of models is that they are often more useful to us when they are inaccurate than when they are accurate. The process of thinking about the failure of a model to explain one or more cases can generate a new causal theory. Glaring inaccuracies often point us in the direction of fruitful theoretical progress.

1.5 RULES OF THE ROAD TO SCIENTIFIC KNOWLEDGE ABOUT POLITICS

In the chapters that follow, we will focus on particular tools of political science research. As we do this, try to keep in mind our larger purpose – trying to advance the state of scientific knowledge about politics. As scientists, we have a number of basic rules that should never be far from our thinking:

- Make your theories causal.
- Don't let data alone drive your theories.
- Consider only empirical evidence.
- Avoid normative statements.
- Pursue both generality and parsimony.

1.5.1 Make Your Theories Causal

All of Chapter 3 deals with the issue of causality and, specifically, how we identify causal relationships. When political scientists construct theories, it is critical that they always think in terms of the causal processes that drive the phenomena in which they are interested. For us to develop a better understanding of the political world, we need to think in terms of causes and not mere covariation. The term covariation is used to describe a situation in which two variables vary together (or covary). If we imagine two variables, A and B, then we would say that A and B covary if it is the case that, when we observe higher values of variable A, we generally also observe higher values of variable B. We would also say that A and B covary if it is the case that, when we observe higher values of variable A, we generally also observe lower values of variable B.[5] It is easy to assume that when we observe covariation we are also observing causality, but it is important not to fall into this trap.

1.5.2 Don't Let Data Alone Drive Your Theories

This rule of the road is closely linked to the first. A longer way of stating it is "try to develop theories before examining the data on which you will perform your tests." The importance of this rule is best illustrated by a silly example. Suppose that we are looking at data on the murder rate (number of murders per 1000 people) in the city of Houston, Texas, by months of the year. This is our dependent variable, and we want to explain why it is higher in some months and lower in others. If we were to take as many different independent variables as possible and simply see whether they had a relationship with our dependent variable, one variable that we might find to strongly covary with the murder rate is the amount of money spent per capita on ice cream. If we perform some verbal gymnastics, we might develop a "theory" about how heightened blood sugar levels in people who eat too much ice cream lead to murderous patterns of behavior. Of course, if we think about it further, we might realize that both ice cream sales and the number of murders committed go up when temperatures rise. Do we have a causally plausible explanation for why temperatures and murder rates might be causally related? It is pretty well known that people's tempers tend to fray when the temperature is higher. People also spend a lot more time outside during hotter weather, and these two factors might combine to produce a plausible relationship between temperatures and murder rates.

[5] A closely related term is correlation. For now we use these two terms interchangeably. In Chapter 8, you will see that there are precise statistical measures of covariance and correlation that are closely related to each other but produce different numbers for the same data.

What this rather silly example illustrates is that we don't want our theories to be crafted based entirely on observations from real-world data. We are likely to be somewhat familiar with empirical patterns relating to the dependent variables for which we are developing causal theories. This is normal; we wouldn't be able to develop theories about phenomena about which we know nothing. But we need to be careful about how much we let what we see guide our development of our theories. One of the best ways to do this is to think about the underlying causal process as we develop our theories and to let this have much more influence on our thinking than patterns that we might have observed.

1.5.3 Consider Only Empirical Evidence

As we previously outlined, we need to always remain open to the possibility that new evidence will come along that will decrease our confidence in even a well-established theory. A closely related rule of the road is that, as scientists, we want to base what we know on what we see from *empirical* evidence, which, as we have said, is simply "evidence based on observing the real world." Strong logical arguments are a good start in favor of a theory, but before we can be convinced, we need to see results from rigorous hypothesis tests.[6]

1.5.4 Avoid Normative Statements

Normative statements are statements about how the world ought to be. Whereas politicians make and break their political careers with normative statements, political scientists need to avoid them at all costs. Most political scientists care about political issues and have opinions about how the world ought to be. On its own, this is not a problem. But when normative preferences about how the world "should" be structured creep into their scientific work, the results can become highly problematic. The best way to avoid such problems is to conduct research and report your findings in

[6] It is worth noting that some political scientists use data drawn from experimental settings to test their hypotheses. There is some debate about whether such data are, strictly speaking, empirical or not. We discuss political science experiments and their limitations in Chapter 4. In recent years some political scientists have also made clever use of simulated data to gain leverage on their phenomena of interest, and the empirical nature of such data can certainly be debated. In the context of this textbook we are not interested in weighing in on these debates about exactly what is and is not empirical data. Instead, we suggest that one should always consider the overall quality of data on which hypothesis tests have been performed when evaluating causal claims.

such a fashion that it is impossible for the reader to tell what your values are or your normative preferences about the world are.

This does not mean that good political science research cannot be used to change the world. To the contrary, advances in our scientific knowledge about phenomena enable policy makers to bring about changes in an effective manner. For instance, if we want to rid the world of wars (normative), we need to understand the systematic dynamics of the international system that produce wars in the first place (empirical and causal). If we want to rid America of homelessness (normative), we need to understand the pathways into and out of being homeless (empirical and causal). If we want to help our favored candidate win elections (normative), we need to understand what characteristics make people vote the way they do (empirical and causal).

1.5.5 Pursue Both Generality and Parsimony

Our final rule of the road is that we should always pursue generality and parsimony. These two goals can come into conflict. By "generality," we mean that we want our theories to be applied to as general a class of phenomena as possible. For instance, a theory that explains the causes of a phenomenon in only one country is less useful than a theory that explains the same phenomenon across multiple countries. Additionally, the more simple or parsimonious a theory is, the more appealing it becomes.[7]

In the real world, however, we often face trade-offs between generality and parsimony. This is the case because, to make a theory apply more generally, we need to add caveats. The more caveats that we add to a theory, the less parsimonious it becomes.

1.6 A QUICK LOOK AHEAD

You now know the rules of the road. As we go through the next 11 chapters, you will acquire an increasingly complicated set of tools for developing and testing scientific theories about politics, so it is crucial that, at every step along the way, you keep these rules in the back of your mind. The rest of this book can be divided into three different sections. The first section, which includes this chapter through Chapter 4, is focused on the development

[7] The term "parsimonious" is often used in a relative sense. So, if we are comparing two theories, the theory that is simpler would be the more parsimonious. Indeed, this rule of the road might be phrased "pursue both generality and simplicity." We use the words "parsimony" and "parsimonious" because they are widely used to describe theories.

of theories and research designs to study causal relationships about politics. In Chapter 2, "The Art of Theory Building," we discuss a range of strategies for developing theories about political phenomena. In Chapter 3, "Evaluating Causal Relationships," we provide a detailed explanation of the logic for evaluating causal claims about relationships between an independent variable, which we call "X," and a dependent variable, which we call "Y." In Chapter 4, "Research Design," we discuss the research strategies that political scientists use to investigate causal relationships.

In the second section of this book, we expand on the basic tools that political scientists need to test their theories. Chapter 5, "Measurement," is a detailed discussion of how we measure (or operationalize) our variables. Chapter 6, "Descriptive Statistics and Graphs," introduces a set of tools that can be used to summarize the characteristics of variables one at a time. Chapter 7, "Statistical Inference," is an introduction to the logic of statistical hypothesis testing. In Chapter 8, "Bivariate Hypothesis Testing," we begin to apply the lessons from Chapter 7 to a series of empirical tests of the relationship between pairs of variables.

The third and final section of this book introduces the critical concepts of the regression model. Chapter 9, "Bivariate Regression Models," introduces the two-variable regression model as an *extension of the concepts* from Chapter 8. In Chapter 10, "Multiple Regression Models I: The Basics," we introduce the multivariate-regression model, with which researchers are able to look at the effects of independent variable X on dependent variable Y while controlling for the effects of other independent variables. Chapter 11, "Multiple Regression Models II: Crucial Extensions," and Chapter 12, "Multiple Regression Models III: Applications," provide in-depth *discussions of* and *advice for* commonly encountered research scenarios involving multivariate-regression models.

CONCEPTS INTRODUCED IN THIS CHAPTER

causal
correlation
covary (or covariation)
dependent variable
empirical
hypothesis
hypothesis testing
independent variable
normal science
normative statements
null hypothesis
operationalize
paradigm
paradigm shift
parsimonious
theoretical models
theory
variable
variable label
variable values

EXERCISES

1. Think about something in the political world that you would like to to better understand. Try to think about this as a variable with high and low values. This is your dependent variable at the conceptual level. Now think about what might cause the values of your dependent variable to be higher or lower. Try to phrase this in terms of an independent variable, also at the conceptual level. Write a paragraph about these two variables and your theory about why they are causally related to each other.
2. Identify something in the world that you would like to see happen (normative). What scientific knowledge (empirical and causal) would help you to pursue this goal?
3. The 1992 U.S. presidential election, in which challenger Bill Clinton defeated incumbent George H. W. Bush, has often been remembered as the "It's the economy, stupid," election. How can we restate the causal statement that embodies this conventional wisdom – "Clinton beat Bush because the economy had performed poorly" – into a more general theoretical statement?

   For Exercises 4 and 5, consider the following statement about the world: "If you care about economic success in a country, you should also care about the peoples' political rights in that country. In a society in which people have more political rights, the victims of corrupt business practices will work through the system to get things corrected. As a result, countries in which people have more political rights will have less corruption. In countries in which there is less corruption, there will be more economic investment and more economic success."
4. Identify at least two causal claims that have been made in the preceding statement. For each causal claim, identify which variable is the independent variable and which variable is the dependent variable. These causal claims should be stated in terms of one of the following types of phrases in which the first blank should be filled by the independent variable and the second blank should be filled by the dependent variable:

   ________________ **causes** ________________

   higher ________________ **causes lower** ________________

   higher ________________ **causes higher** ________________

5. Draw a graph like Figure 1.3 for each of the causal claims that you identified in Exercise 4. For each of your figures, do the following: Start on the left-hand side of the horizontal axis of the figure. This should represent a low value of the independent variable. What value of the dependent variable would you expect to find for such a case? Put a dot on your figure that represents this expected location. Now do the same for a case with a high value of the independent variable. Draw a line that connects these two points and write a couple of sentences that describe this picture.

6. Find an article in a political science journal that contains a model of politics. Provide the citation to the article, and answer the following questions:
   (a) What is the dependent variable?
   (b) What is one of the independent variables?
   (c) What is the causal theory that connects the independent variable to the dependent variable?
   (d) Does this seem reasonable?

2 The Art of Theory Building

OVERVIEW

In this chapter we discuss the art of theory building. Unfortunately there is no magical formula or cookbook for developing good theories about politics. But there are strategies for developing theories that will help you to develop good theories. We discuss these strategies in this chapter.

2.1 GOOD THEORIES COME FROM GOOD THEORY-BUILDING STRATEGIES

In Chapter 1 we discussed the role of theories in developing scientific knowledge. From that discussion, it is clear that a "good" theory is one that, after going through the rigors of the evaluation process, makes a contribution to scientific knowledge. In other words, a good theory is one that changes the way that we think about some aspect of the political world. We also know from our discussion of the rules of the road that we want our theories to be causal, empirical, nonnormative, general, and parsimonious. This is a tall order, and a logical question to ask at this point is "How do I come up with such a theory?"

Unfortunately, there is neither an easy answer nor a single answer. Instead, what we can offer you is a set of strategies. "Strategies?" you may ask. Imagine that you were given the following assignment: "Go out and get struck by lightning."[1] There is no cut-and-dried formula that will show you how to get struck by lightning, but certainly there are actions that you can take that will make it more likely. The first step is to look at a weather map and find an area where there is thunderstorm activity, and if you were to go to such an area, you would increase your likelihood of getting struck.

[1] Our lawyers have asked us to make clear that this is an illustrative analogy and that we are in no way encouraging you to go out and try to get struck by lightning.

You would be even more likely to get struck by lightning if, once in the area of thunderstorms, you climbed to the top of a tall barren hill. But you would be still more likely to get struck if you carried with you a nine iron and, once on top of the barren hill, in the middle of a thunderstorm, you held that nine iron up to the sky. The point here is that, although there are no magical formulae that make the development of a good theory (or getting hit by lightning) a certain event, there are strategies that you can follow to increase the likelihood of it happening.

2.2 IDENTIFYING INTERESTING VARIATION

A useful first step in theory building is to think about phenomena that vary and to focus on general patterns. Because theories are designed to explain variation in the dependent variable, identifying some variation that is of interest to you is a good jumping-off point. In Chapter 4 we present a discussion of two of the most common research designs – cross-sectional and time-series observational studies – in some detail. For now it is useful to give a brief description of each in terms of the types of variation in the dependent variable. These should help clarify the types of variation to consider as you begin to think about potential research ideas.

When we think about measuring our dependent variable, the first things that we need to identify are the time and spatial dimensions over which we would like to measure this variable. The **time dimension** identifies the point or points in time at which we would like to measure our variable. Depending on what we are measuring, typical time increments for political science data are annual, quarterly, monthly, or weekly measures. The **spatial dimension** identifies the units that we want to measure. There is a lot of variability in terms of the spatial units in political science data. If we are looking at survey data, the spatial unit will be the individual people who answered the survey (known as survey respondents). If we are looking at data on U.S. state governments, the typical spatial unit will be the 50 U.S. states. Data from international relations and comparative politics often take nations as their spatial units. Throughout this book, we think about measuring our dependent variable such that one of these two dimensions will be static (or constant). This means that our measures of our dependent variable will be of one of two types. The first is a **time-series measure**, in which the spatial dimension is the same for all cases and the dependent variable is measured at multiple points in time. The second is a **cross-sectional measure**, in which the time dimension is the same for all cases and the dependent variable is measured for multiple spatial units. Although it is possible for us to measure the same variable across both time and space, we strongly recommend thinking in terms of variation across only one of these

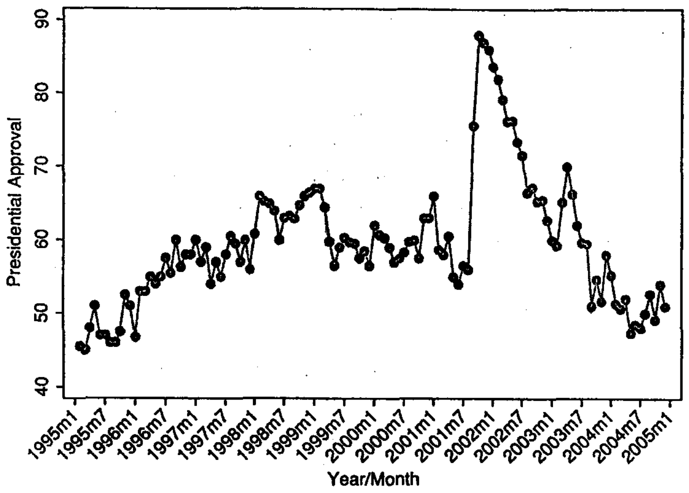

Figure 2.1. Presidential approval, 1995–2005.

two dimensions as you attempt to develop a theory about what causes this variation.[2] Let's consider an example of each type of dependent variable.

2.2.1 Time-Series Example

In Figure 2.1 we see the average monthly level of U.S. presidential approval displayed from 1995 to 2005. We can tell that this variable is measured as a time series because the spatial unit is the same (the United States), but the variable has been measured at multiple points in time (each month). This measure is comparable across the cases; for each month we are looking at the average percentage of people who reported that they approved of the job that the president was doing. Once we have a measure like this that is comparable across cases, we can start to think about what independent variable might *cause* the level of the dependent variable to be higher or lower.

If you just had a mental alarm bell go off telling you that we seemed to be violating one of our rules of the road from Chapter 1, then congratulations – you are doing a good job paying attention. Our second rule of the road is "don't let data alone drive your theories." Remember that we

[2] As we mentioned in Chapter 1, we will eventually theorize about multiple independent variables simultaneously causing the same dependent variable to vary. Confining variation in the dependent variable to a single dimension helps to make such multivariate considerations tractable.

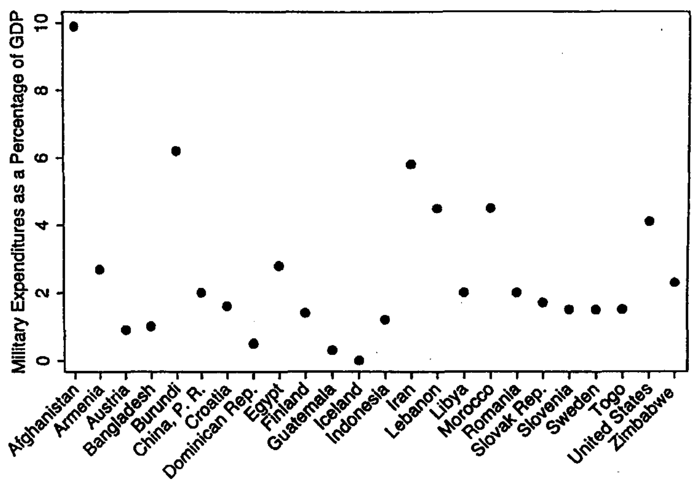

Figure 2.2. Military spending in 2005.

also can phrase this rule as "try to develop theories before examining the data on which you will perform your tests." So what this means is that we might develop a theory about U.S. presidential approval using Figure 2.1, but we would want to test that theory by using a different set of data that may or may not contain the data depicted in Figure 2.1.

2.2.2 Cross-Sectional Example

In Figure 2.2 we see military spending as a percentage of gross domestic product (GDP) in 2005 for 24 randomly selected nations. We can tell that this variable is measured cross sectionally, because it varies across spatial units (nations) but does not vary across time (it is measured for the year 2005 for each case). When we measure variables across spatial units like this, we have to be careful to choose appropriate measures that are comparable across spatial units. To better understand this, imagine that we had measured our dependent variable as the amount of money that each nation spent on its military. The problem would be that country currencies – the Afghan afghani, the Armenian dram, and Austrian euro – do not take on the same value. We would need to know the currency exchange rates in order to make these comparable across nations. Using currency exchange rates, we would be able to convert the absolute amounts of money that each nation had spent into a common measure. We could think of this particular measure as an operationalization of the concept of relative military "might." This would be a perfectly reasonable dependent variable for theories about what makes one nation more powerful than another. Why, you might ask, would we want to measure military spending as a percentage of GDP? The answer is that this comparison is our attempt to measure the percentage of the total budgetary effort available that a nation is putting into its armed forces. Some nations have larger economies than others, and this measure allows us to answer the question of how much of their total economic activity each nation is putting toward its military. We can theorize about what would *cause* a nation to put more or less of its available economic resources toward military spending.

Of course, as we discussed in the previous subsection, we would not want to develop our theory by using data from these 24 cases and then test it by using only the same set of cases.

2.3 LEARNING TO USE YOUR KNOWLEDGE

One of the common problems that people have when trying to develop a theory about a phenomenon of interest is that they can't get past a particular political event in time or a particular place about which they know a lot. It is helpful to know some specifics about politics, but it is also important to be able to distance yourself from the specifics of one case and to think more *broadly* about the underlying causal process. To use an analogy, it's fine to know something about trees, but we want to theorize about the forest. Remember, one of our rules of the road is to try to make our theories general.

2.3.1 Moving from a Specific Event to More General Theories

For an example of this, return to Figure 2.1. What is the first thing that you think most people notice when they look at Figure 2.1? Once they have figured out what the dimensions are in this figure (U.S. presidential approval over time), many people look at the fall of 2001 and notice the sharp increase in presidential approval that followed the terrorist attacks on the United States on September 11, 2001. This is a period of recent history about which many people have detailed memories. In particular, they might remember how the nation rallied around President Bush in the aftermath of these attacks. There are few people who would doubt that there was a causal linkage between these terrorist attacks and the subsequent spike in presidential approval.

At first glance, this particular incident might strike us as a unique event from which general theoretical insights cannot be drawn. After all, terrorist attacks on U.S. soil are rare events, and attacks of this magnitude are even more rare. The challenge to the scientific mind when we have strong

confidence about a causal relationship in one specific incident is to push the core concepts around in what we might call thought experiments: How might a less-effective terrorist attack affect public opinion? How might other types of international incidents shape public opinion? Do we think that terrorist attacks lead to similar reactions in public opinion toward leaders in other nations? Each of these questions is posed in general terms, taking the specific events of this one incident as a jumping-off point. The answers to these more general questions should lead us to general theories about the causal impact of international incidents on public opinion.

In the 1970s John Mueller moved from the specifics of particular international incidents and their influence on public opinion towards a general theory of what causes rallies (or short-term increases) in public opinion.[3] Mueller developed a theory that presidential popularity would increase in the short term any time that there was international conflict. Mueller thought that this would occur because, in the face of international conflict, people would tend to put their partisan differences and other critiques that they may have of the president's handling of his job aside and support him as the commander and chief of the nation. In Mueller's statistical analysis of time-series data on presidential approval, he found that there was substantial support for his hypothesis that international conflicts would raise presidential approval rates, and this in turn gave him confidence in his theory of public opinion rallies.

2.3.2 Know Local, Think Global: Can You Drop the Proper Nouns?

Physicists don't have theories that apply only in France, and neither should we. Yet many political scientists write articles with one particular geographic context in mind. Among these, the articles that have the greatest impact are those that advance general theories from which the proper nouns have been removed.[4] An excellent example of this is Michael Lewis-Beck's article titled "Who's the Chef?" Lewis-Beck, like many observers of French politics, had observed the particularly colorful period from 1986 to 1988 during which the president was a socialist named François Mitterand and the prime minister was Jacques Chirac, a right-wing politician from the Guallist RPR party. The height of this political melodrama occurred when both leaders showed up to international summits of world leaders claiming to be the rightful representative of the French Republic. This led to a famous photo of the leaders of the G-7 group of nations that contained eight

[3] See Mueller (1973).

[4] By "proper nouns," we mean specific names of people or countries. But this logic can and should be pushed further to include specific dates, as we subsequently argue.

people.[5] Although many people saw this as just another colorful anecdote about the ever-changing nature of the power relationship between presidents and prime ministers in Fifth Republic France, Lewis-Beck moved from the specifics of such events to develop and test a general theory about political control and public opinion.

His theory was that changing the political control of the economy would cause public opinion to shift in terms of who was held accountable for the economy. In France, during times of unified political control of the top offices, the president is dominant, and thus according to Lewis-Beck's theory the president should be held accountable for economic outcomes. However, during periods of divided control, Lewis-Beck's theory leads to the expectation that the prime minister, because of his or her control of economic management during such periods, should be held accountable for economic outcomes. Through careful analysis of time-series data on political control and economic accountability, Lewis-Beck found that his theory was indeed supported.

Although the results of this study are important for advancing our understanding of French politics, the theoretical contribution made by Lewis-Beck was much greater because he couched it in general terms and without proper nouns. We also can use this logic to move from an understanding of a specific event to general theories that explain variation across multiple events. For example, although it might be tempting to think that every U.S. presidential election is entirely unique – with different candidates (proper names) and different historical circumstances – the better scientific theory does *not* explain only the outcome of the 2008 U.S. presidential election, but of U.S. presidential elections in general. That is, instead of asking "Why did Bush beat Kerry in the 2004 election?" we should ask either "What causes incumbent success rates in U.S. presidential elections?" or "What causes Republican candidates to fare better or worse than Democratic candidates in U.S. presidential elections?"

2.4 EXAMINE PREVIOUS RESEARCH

Once you have identified an area in which you want to conduct research, it is often useful to look at what other work has been done that is related to your areas of interest. As we discussed in Chapter 1, part of taking a scientific approach is to be skeptical of research findings, whether they are our own or those of other researchers. By taking a skeptical look at the

[5] The G-7, now the G-8 with the inclusion of Russia, is an annual summit meeting of the heads of government from the world's most powerful nations.

research of others, we can develop new research ideas of our own and thus develop new theories.

We therefore suggest looking at research that seems interesting to you and, as you examine what has been done, keep the following list of questions in mind:

- What (if any) other causes of the dependent variable did the previous researchers miss?
- Can their theory be applied elsewhere?
- If we believe their findings, are there further implications?
- How might this theory work at different levels of aggregation (micro$\Longleftrightarrow$macro)?

2.4.1 What Did the Previous Researchers Miss?

Any time that we read the work of others, the first thing that we should do is break down their theory or theories in terms of the independent and dependent variables that they claim are causally related to each other.[6] Once we have done this, we should think about whether the causal arguments that other researchers have advanced seem reasonable. (In Chapter 3 we present a detailed four-step process for doing this.) We should also be in the habit of coming up with other independent variables that we think might be causally related to the same dependent variable. Going through this type of mental exercise can lead to new theories that are worth pursuing.

2.4.2 Can Their Theory Be Applied Elsewhere?

When we read about the empirical research that others have conducted, we should be sure that we understand which specific cases they were studying when they tested their theory. We should then proceed with a mental exercise in which we think about what we might find if we applied the same theory to other cases. In doing so, we will probably identify some cases for which we expect to get the same results, as well as other cases for which we might have different expectations. Of course, we would have to carry out our own empirical research to know whether our speculation along these lines is correct, but replicating research can lead to interesting findings. The most useful theoretical development comes when we can identify systematic patterns in the types of cases that will fit and those that will not fit the

[6] We cannot overstate the importance of this endeavor. We understand that this can be a difficult task for a beginning student, but it gets easier with practice. A good way to start this process is to look at the figures or tables in an article and ask yourself, "What is the dependent variable here?"

established theory. These systematic patterns are additional variables that determine whether a theory will work across an expanded set of cases. In this way we can think about developing new theories that will subsume the original established theory.

2.4.3 If We Believe Their Findings, Are There Further Implications?

Beginning researchers often find themselves intimidated when they read convincing accounts of the research carried out by more established scholars. After all, how can we ever expect to produce such innovative theories and find such convincingly supportive results from extensive empirical tests? Instead of being intimidated by such works, we need to learn to view them as opportunities – opportunities to carry their logic further and think about what other implications might be out there. If, for example, another researcher has produced a convincing theory about how voters behave, we could ask how might this new understanding alter the behavior of strategic politicians who understand that voters behave in this fashion?

One of the best examples of this type of research extension in political science comes from our previous example of John Mueller's research on rallies in presidential popularity. Because Mueller had found such convincingly supportive evidence of this "rally 'round the flag effect" in his empirical testing, other researchers were able to think through the strategic consequences of this phenomenon. This led to a new body of research on a phenomenon called "diversionary use of force." The idea of this new research is that, because strategic politicians will be aware that international conflicts temporarily increase presidential popularity, they will choose to generate international conflicts at times when they need such a boost.

2.4.4 How Might This Theory Work at Different Levels of Aggregation (Micro⟺Macro)?

As a final way to use the research of others to generate new theories we suggest considering how a theory might work differently at varying levels of aggregation. In political science research, the lowest level of aggregation is usually at the level of individual people in studies of public opinion. As we saw in Subsection 2.4.3, when we find a trend in terms of individual-level behavior, we can develop new theoretical insights by thinking about how strategic politicians might take advantage of such trends. Sometimes it is possible to gain these insights by simply changing the level of aggregation. As we have seen, political scientists have often studied trends in public opinion by examining data measured at the national level over time. This type of study is referred to as the study of macro politics. When we

find trends in public opinion at higher (macro) levels of aggregation, it is always an interesting thought exercise to consider what types of patterns of individual-level or "micro-" level behavior are driving these aggregate-level findings.

As an example of this, return to the rally 'round the flag example and change the level of aggregation. We have evidence that, when there are international conflicts, public opinion toward the president becomes more positive. What types of individual-level forces might be driving this observed aggregate-level trend? It might be the case that there is a uniform shift across all types of individuals in their feelings about the president. It might also be the case that the shift is less uniform. Perhaps individuals who dislike the president's policy positions on domestic events are willing to put these differences aside in the face of international conflicts, whereas the opinions of the people who were already supporters of the president remain unchanged. Thinking about the individual-level dynamics that drive aggregate observations can be a fruitful source of new causal theories.

2.5 THINK FORMALLY ABOUT THE CAUSES THAT LEAD TO VARIATION IN YOUR DEPENDENT VARIABLE

Thus far in this book we have discussed thinking about the political world in an organized, systematic fashion. By now, we hope that you are starting to think about politics in terms of independent variables and dependent variables and are developing theories about *the causal relationships* between them. The theories that we have considered thus far have come from thinking rigorously about the phenomena that we want to explain and deducing plausible causal explanations. One extension of this type of rigorous thinking is labeled "formal theory" or "rational choice."[7]

The formal-theory approach to social science phenomena starts out with a fairly basic set of assumptions about human behavior and then uses game theory and other mathematical tools to build models of phenomena of interest. We can summarize these assumptions about human behavior by saying that formal theorists assume that all individuals are **rational utility maximizers** – that they attempt to maximize their self-interest. Individuals are faced with a variety of choices in political interactions, and those choices carry with them different consequences – some desirable, others

[7] The terms "formal theory" and "rational choice" have been used fairly interchangeably to describe the application of game theory and other formal mathematical tools to puzzles of human behavior. We have a slight preference for the term "formal theory" because it is a more overarching term describing the enterprise of using these tools, whereas "rational choice" describes the most critical assumption that this approach makes.

undesirable. By thinking through the incentives faced by individuals, users of this approach begin with the strategic foundations of the decisions that individuals face. Formal theorists then deduce theoretical expectations of what individuals will do given their preferences and the strategic environment that they confront.

That sounds like a mouthful, we know. Let's begin with a simple example: If human beings are self-interested, then (by definition) members of a legislature are self-interested. This assumption suggests that members will place a high premium on reelection. Why is that? Because, first and foremost, a politician must be in office if she is going to achieve her political goals. And from this simple deduction flows a whole set of hypotheses about congressional organization and behavior.[8]

This approach to studying politics is a mathematically rigorous attempt to think through what it would be like to be in the place of different actors involved in a situation in which they have to choose how to act. In essence, formal theory is a lot like the saying that we should not judge a person until we have walked a mile in his or her shoes. We use the tools of formal theory to try to put ourselves in the position of imagining that we are in someone else's shoes and thinking about the different choices that he or she has to make. In the following subsections we introduce the basic tools for doing this by using an expected utility approach and then provide a famous example of how researchers used this framework to develop theories about why people vote.

2.5.1 Utility and Expected Utility

Think about the choice that you have made to read this chapter of this book. What are your expected benefits and what are the costs that you expect to incur? One benefit may be that you are genuinely curious about how we build theories of politics. Another expected benefit may be that your professor is likely to test you on this material, and you expect that you will perform better if you have read this chapter. There are, no doubt, also costs to reading this book. What else might you be doing with your time? This is the way that formal theorists approach the world.

Formal theorists think about the world in terms of the outcome of a collection of individual-level decisions about what to do. In thinking about an individual's choices of actions, formal theorists put everything in terms of utility. Utility is an intentionally vague quantity. The utility from a particular action is equal to the sum of all benefits minus the sum of all costs from that action. If we consider an action Y, we can summarize the

[8] See Mayhew (1974) and Fiorina (1989).

utility from Y for individual i with the following formula:

$$U_i(Y) = \sum B_i(Y) - \sum C_i(Y),$$

where $U_i(Y)$ is the utility for individual i from action Y, $\sum B_i(Y)$ is the sum of the benefits B_i from action Y for individual i, and $\sum C_i(Y)$ is the sum of the costs C_i from action Y for individual i. When choosing among a set of possible actions (including the decisions not to act), a rational individual will choose that action that maximizes their utility. To put this formally,

$$\text{given a set of choices } Y = Y_1, Y_2, Y_3, \ldots, Y_n,$$

$$\text{individual } i \text{ will choose } Y_a \text{ such that } U_i(Y_a) > U_i(Y_b) \; \forall \; b \neq a,$$

which translates into, "given a set of choices of action Y_1 through Y_n, individual i will choose that action (Y_a) such that the utility to individual i from that action is greater than the utility to individual i from any action (Y_b) for all ($\forall$) actions b not equal to a." In more straightforward terms, we could translate this into the individual choosing that action which he deems best for himself.

At this point, it is reasonable to look around the real world and think about exceptions. Is this really the way that the world works? What about altruism? During the summer of 2006, the world's second-richest man, Warren Buffet, agreed to donate more than 30 billion dollars to the Bill and Melinda Gates Foundation. Could this possibly have been rational utility-maximizing behavior? What about suicide bombers? The answer to these types of questions shows both the flexibility and a potential problem of the concept of utility. Note that, in the preceding formulae, there is always a subscripted i under each of the referenced utility components, (U_i, B_i, C_i). This is because different individuals have *different* evaluations of the benefits (B_i) and costs (C_i) associated with a particular action. When the critic of this approach says, "How can this possibly be utility-maximizing behavior?" the formal theorist responds, "Because this is just an individual with an unusual utility structure."

Think of it another way. Criticizing formal theory because it takes preferences as "given" – that is, as predetermined, rather than the focus of inquiry – strikes us as beside the point. Other parts of political science can and should study preference formation; think about political psychology and the study of public opinion. What formal theory does, and does well, is to say, "Okay, once an individual has her preferences – regardless of where they came from – how do those preferences interact with strategic opportunities and incentives to produce political outcomes?" Because formal theory takes those preferences as given does not mean that the preference-formation process is unimportant. It merely means that formal theory is here to explain a different portion of social reality.

From a scientific perspective, this is fairly unsettling. As we discussed in Chapter 1, we want to build scientific knowledge based on real-world observation. How do we observe people's utilities? Although we can ask people questions about what they like and don't like, and even their perceptions of costs and benefits, we can never truly observe utilities. Instead, the assumption of utility maximization is just that – an assumption. This assumption is, however, a fairly robust assumption, and we can do a lot if we are willing to make it and move forward while keeping the potential problems in the back of our minds.

Another potentially troubling aspect of the rational-actor utility-maximizing assumption that you may have thought of is the assumption of complete information. In other words, what if we don't know exactly what the costs and benefits will be from a particular action? In the preceding formulae, we were operating under the assumption of complete information, for which we knew exactly what would be the costs, benefits, and thus the utility from each possible action. When we relax this assumption, we move our discussion from utility to expected utility. This is a pretty straightforward transformation in which we put expectations in front of all utilities. So, under incomplete information, for an individual action Y,

$$E[U_i(Y)] = \sum E[B_i(Y)] - \sum E[C_i(Y)],$$

and a rational actor will maximize his expected utility thus:

given a set of choices $Y = Y_1, Y_2, Y_3, \ldots, Y_n$,

individual i will choose Y_a such that $E[U_i(Y_a)] > E[U_i(Y_b)] \forall b \neq a$.

2.5.2 The Puzzle of Turnout

One of the oldest and enduring applications of formal theory to politics is known as the "paradox of voting." William Riker and Peter Ordeshook set out the core arguments surrounding this application in their influential 1968 article in the *American Political Science Review* titled "A Theory of the Calculus of Voting." Their paper was written to weigh in on a lively debate over the rationality of voting. In Riker and Ordeshook's notation (with subscripts added), the expected utility of voting was summarized as

$$R_i = (B_i P_i) - C_i,$$

where R_i is the reward that an individual receives from voting, B_i is the differential benefit that an individual voter receives "from the success of his more preferred candidate over his less preferred one" (Riker and Ordeshook 1968, p. 25), P_i is the probability that that voter will cast the deciding vote that makes her preferred candidate the winner, and C_i is the sum of the costs that the voter incurs from voting.[9] If R_i is positive, the individual votes; otherwise, she abstains.

We'll work our way through the right-hand side of this formula and think about the likely values of each term in this equation for an individual eligible voter in a U.S. presidential election. The term B_i is likely to be greater than zero for most eligible voters in most U.S. presidential elections. The reasons for this vary widely from policy preferences to gut feelings about the relative character traits of the different candidates. Note, however, that the B_i term is multiplied by the P_i term. What is the likely value of P_i? Most observers of elections would argue that P_i is extremely small and effectively equal to zero for every voter in most elections. In the case of a U.S. presidential election, for one vote to be decisive, that voter must live in a state in which the popular vote total would be *exactly* tied, and this must be a presidential election for which that particular state would swing the outcome in the Electoral College to either candidate. Because P_i is effectively equal to zero, the entire term $(B_i P_i)$ is effectively equal to zero.

What about the costs of voting, C_i? Voting takes time for all voters. Even if a voter lives right next door to the polling place, she has to take some time to walk next door, perhaps stand in a line, and cast her ballot. The well-worn phrase "time is money" certainly applies here. Even if the voter in question is not working at the time that she votes, she could be doing something other than voting. Thus it is pretty clear that C_i is greater than zero. If C_i is greater than zero and $(B_i P_i)$ is effectively equal to zero, then R_i must be negative. How then, do we explain the millions of people that vote in U.S. presidential elections, or, indeed, elections around the world? Is this evidence that people are truly not rational? Or, perhaps, is it evidence that millions of people systematically overestimate P_i? Influential political economy scholars, including Anthony Downs and Gordon Tullock, posed these questions in the early years of formal theoretical analyses of politics.

Riker and Ordeshook's answer was that there must be some other benefit to voting that is not captured by the term $(B_i P_i)$. They proposed that the voting equation should be

$$R_i = (B_i P_i) - C_i + D_i,$$

[9] For simplicity in this example, consider an election in which there are only two candidates competing. Adding more candidates makes the calculation of B_i more complicated, but does not change the basic result of this model.

where D_i is the satisfaction that individuals feel from participating in the democratic process, regardless of the impact of their participation on the final outcome of the election. Riker and Ordeshook argued that D_i could be made up of a variety of different efficacious feelings about the political system, ranging from fulfilling one's duties as a citizen to standing up and being counted.

Think of the contribution that Riker and Ordeshook made to political science, and that, more broadly, formal theory makes to political science, in the following way: Riker and Ordeshook's theory leads us to wonder why any individual will vote. And yet, empirically, we notice that close to half of the adult population votes in any given presidential election in recent history. What formal theory accomplishes for us is that it helps us to focus in on exactly *why* people do bother, rather than to assert, normatively, that people *should*.[10]

2.6 THINK ABOUT THE INSTITUTIONS: THE RULES USUALLY MATTER

One rich source for theoretical insights comes from thinking about institutional arrangements and the influence that they have in shaping political behavior and outcomes. In other words, take some time to think about the rules under which the political game is played. To fully understand these rules and their impact, we need to think through some counterfactual scenarios in which we imagine how outcomes would be altered if there were different rules in place. This type of exercise can lead to some valuable theoretical insights. In the subsections that follow, we consider two examples of thinking about the impact of institutions.

2.6.1 Legislative Rules

Considering the rules of the political game has yielded theoretical insights into the study of legislatures and other governmental decision-making

[10] Of course, Riker and Ordeshook did not have the final word in 1968. In fact, the debate over the rationality of turnout has been at the core of the debate over the usefulness of formal theory in general. In their 1994 book titled *Pathologies of Rational Choice Theory*, Donald Green and Ian Shapiro made the first point of attack in their critique of the role that formal theory plays in political science. One of Green and Shapiro's major criticisms of this part of political science was that the linkages between formal theory and empirical hypothesis tests were too weak. In reaction to these and other critics, the National Science Foundation launched a new program titled "Empirical Implications of Theoretical Models" (EITM) that was designed to strengthen the linkage between formal theory and empirical hypothesis tests.

bodies. This has typically involved thinking about the preference orderings of expected utility-maximizing actors. For example, let's imagine a legislature made up of three individual members, X, Y, and Z.[11] The task in front of X, Y, and Z is to choose between three alternatives A, B, and C. The preferences orderings for these three rational individuals are as follows:

$$X : ABC,$$

$$Y : BCA,$$

$$Z : CAB.$$

An additional assumption that is made under these circumstances is that the preferences of rational individuals are **transitive**. This means that if individual X likes A better than B and B better than C, then, for X's preferences to be transitive, he or she must also like A better than C. Why is this an important assumption to make? Consider the alternative. What if X liked A better than B and B better than C, but liked C better than A? Under these circumstances, it would be impossible to discuss what X wants in a meaningful fashion because X's preferences would produce an infinite cycle. To put this another way, no matter which of the three choices X chose, there would always be some other choice that X prefers. Under these circumstances, X could not make a rational choice.

In this scenario, what would the group prefer? This is not an easy question to answer. If they each voted for their first choice, each alternative would receive one vote. If these three individuals vote between pairs of alternatives, and they vote according to their preferences, we would observe the following results:

$$A \text{ vs. } B,\ X\&Z \text{ vs. } Y,\ A \text{ wins};$$

$$B \text{ vs. } C,\ X\&Y \text{ vs. } Z,\ B \text{ wins};$$

$$C \text{ vs. } A,\ Y\&Z \text{ vs. } X,\ C \text{ wins}.$$

Which of these three alternatives does the group collectively prefer? This is an impossible question to answer because the group's preferences cycle across the three alternatives. Another way of describing this group's preferences is to say that they are **intransitive** (despite the fact that, as you can see, each individual's preferences are transitive).

[11] We know that, in practice, legislatures tend to have many more members. Starting with this type of miniature-scaled legislature makes formal considerations much easier to carry out. Once we have arrived at conclusions based on calculations made on such a small scale, it is important to consider whether the conclusions that we have drawn would apply to more realistically larger-scaled scenarios.

This result should be fairly troubling to people who are concerned with the fairness of democratic elections. One of the often-stated goals of elections is to "let the people speak." Yet, as we have just seen, it is possible that, even when the people involved are all rational actors, their collective preferences may not be rational. Under such circumstances, a lot of the normative concepts concerning the role of elections simply break down. This finding is at the heart of Arrow's theorem, which was developed by Kenneth Arrow in his 1951 book titled *Social Choice and Individual Values*. At the time of its publication, political scientists largely ignored this book. As formal theory became more popular in political science, Arrow's mathematical approach to these issues became increasingly recognized. In 1982 William Riker popularized Arrow's theorem in his book *Liberalism Against Populism*, in which he presented a more accessible version of Arrow's theorem and bolstered a number of Arrow's claims through mathematical expositions.

2.6.2 The Rules Matter!

Continuing to work with our example of three individuals, *X*, *Y*, and *Z*, with the previously described preferences, now imagine that the three individuals will choose among the alternatives in two different rounds of votes between pairs of choices. In the first round of voting, two of the alternatives will be pitted against each other. In the second round of voting, the alternative that won the first vote will be pitted against the alternative that was not among the choices in the first round. The winner of the second round of voting is the overall winning choice.

In our initial consideration of this scenario, we will assume that *X*, *Y*, and *Z* will vote according to their preferences. What if *X* got to decide on the order in which the alternatives got chosen? We know that *X*'s preference ordering is *ABC*. Can *X* set things up so that *A* will win? What if *X* made the following rules:

1st round: *B* vs. *C*;

2nd round: 1st round winner vs. *A*.

What would happen under these rules? We know that both *X* and *Y* prefer *B* to C, so *B* would win the first round and then would be paired against *A* in the second round. We also know that *X* and *Z* prefer *A* to *B*, so alternative *A* would win and *X* would be happy with this outcome.

Does voting like this occur in the real world? Actually, the answer is "yes." This form of pairwise voting among alternatives is the way that legislatures typically conduct their voting. If we think of individuals *X*, *Y*, and *Z* as being members of a legislature, we can see that whoever controls the ordering of the voting (the rules) has substantial power. To explore these issues further, let's examine the situation of individual *Y*. Remember that *Y*'s preference ordering is *BCA*. So *Y* would be particularly unhappy about the outcome of the voting according to *X*'s rules, because it resulted in *Y*'s least-favorite outcome. But remember that, for our initial consideration, we assumed that *X*, *Y*, and *Z* will vote according to their preferences. If we relax this assumption, what might *Y* do? In the first round of voting, *Y* could cast a strategic vote for *C* against *B*. If both *X* and *Z* continued to vote (sincerely) according to their preferences, then *C* would win the first round. Because we know that both *X* and *Z* prefer *C* to *A*, *C* would win the second round and would be the chosen alternative. Under these circumstances, *Y* would be better off because *Y* prefers alternative *C* to *A*.

From the perspective of members of a legislature, it is clearly better to control the rules than to vote strategically to try to obtain a better outcome. When legislators face reelection, one of the common tactics of their opponents is to point to specific votes in which the incumbent appears to have voted contrary to the preferences of his constituents. It would seem reasonable to expect that legislator *Y* comes from a district with the same or similar preferences to those of *Y*. By casting a strategic vote for *C* over *B*, *Y* was able to obtain a better outcome but created an opportunity for an electoral challenger to tell voters that *Y* had voted against the preferences of his district.

In *Congressmen in Committees*, Richard Fenno's classic study of the U.S. House of Representatives, one of the findings was that the Rules Committee – along with the Ways and Means and the Appropriations Committees – was one of the most requested committee assignments from the individual members of Congress. At first glance, the latter two committees make sense as prominent committees and, indeed, receive much attention in the popular media. By contrast, the Rules Committee very rarely gets any media attention. Members of Congress certainly understand and appreciate the fact that the rules matter, and formal theoretical thought exercises like the preceding one help us to see why this is the case.

2.7 EXTENSIONS

These examples truly represent just the beginning of the uses of formal theory in political science. We have not even introduced two of the more important aspects of formal theory – spatial models and game theory – that are beyond the scope of this discussion. In ways that mirror applications in microeconomics, political scientists have used spatial models to study phenomena such as the placement of political parties along the ideological spectrum, much as economists have used spatial models to study the location of firms in a market. Likewise, game theory utilizes a highly structured sequence of moves by different players to show how any particular actor's utility depends not only on her own choices, but also on the choices made by the other actors. It is easy to see hints about how game theory works in the preceding simple three-actor, two-stage voting examples: *X*'s best vote in the first stage likely depends on which alternative *Y* and *Z* choose to support, and vice versa. Game theory, then, highlights how the strategic choices made in politics are interdependent.

2.8 HOW DO I KNOW IF I HAVE A "GOOD" THEORY?

Once you have gone through some or all of the suggested courses of action for building a theory, a reasonable question to ask is, "How do I know if I have a 'good' theory?" Unfortunately there is not a single succinct way of answering this question. Instead, we suggest that you answer a set of questions about your theory and consider your honest answers to these questions as you try to evaluate the overall quality of our theory. You will notice that some of these questions come directly from the "rules of the road" that we developed in Chapter 1:

- Is your theory causal?
- Can you test your theory on data that you have not yet observed?
- How general is your theory?
- How parsimonious is your theory?
- How new is your theory?
- How nonobvious is your theory?

2.8.1 Is Your Theory Causal?

Remember that our first rule of the road to scientific knowledge about politics is "Make your theories causal." If your answer to the question "Is your theory causal?" is anything other than "yes," then you need to go back to the drawing board until the answer is an emphatic "yes."

As scientists studying politics, we want to know why things happen the way that they happen. As such, we will not be satisfied with mere correlations and we demand causal explanations. We know from Chapter 1 that one way initially to evaluate a particular theory is to think about the causal explanation behind it. The causal explanation behind a theory is the answer to the question "Why do you think that this independent variable is causally related to this dependent variable?" If the answer is reasonable, then you can answer this first question with a "yes."

2.8.2 Can You Test Your Theory on Data That You Have Not Yet Observed?

Our second rule of the road is "Don't let data alone drive your theories," which we restated in a slightly longer form as "Try to develop theories before examining the data on which you will perform your tests." If you have derived your theory from considering a set of empirical data, you need to be careful not to have observed all of the data on which you can test your theory. This can be a somewhat gray area, and only you know whether your theory is entirely data driven and whether you observed all of your testing data before you developed your theory.

2.8.3 How General Is Your Theory?

We could rephrase this question for evaluating your theory as "How widely does your theory apply?" To the extent that your theory is not limited to one particular time period or to one particular spatial unit, it is more general. Answers to this question vary along a continuum – it's not the end of the world to have a fairly specific theory, but, all else being equal, a more general theory is more desirable.

2.8.4 How Parsimonious Is Your Theory?

As with the question in the preceding subsection, answers to this question also vary along a continuum. In fact, it is often the case that we face a trade-off between parsimony and generality. In other words, to make a theory more general, we often have to give up parsimony, and to make a theory more parsimonious, we often have to give up generality. The important thing with both of these desirable aspects of a theory is that we have them in mind as we evaluate our theory. If we can make our theory more general or more parsimonious and without sacrifice, we should do so.

2.8.5 How New Is Your Theory?

At first it might seem that this is a pretty straightforward question to answer. The problem is that we cannot know about all of the work that has been done before our own work in any particular area of research. It also is often the case that we may think our theory is really new, and luckily we have not been able to find any other work that has put forward the same theory on the same political phenomenon. But then we discover a similar theory on a related phenomenon. There is no simple answer to this question. Rather, our scholarly peers usually answer this question of newness for us when they evaluate our work.

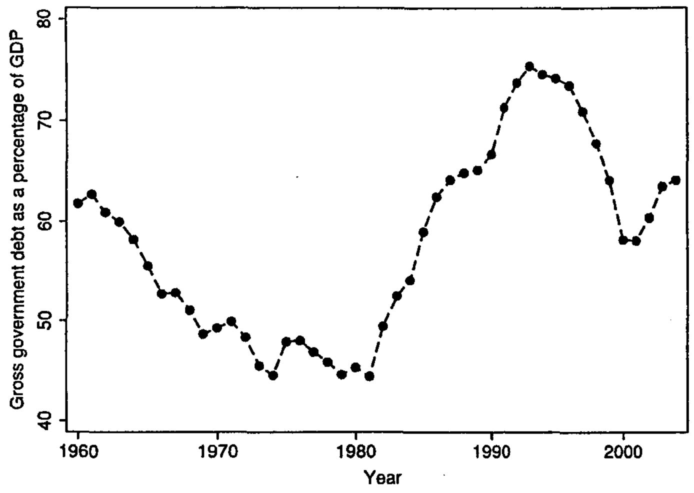

Figure 2.3. Gross U.S. government debt as a percentage of GDP, 1960–2004.

2.8.6 How Nonobvious Is Your Theory?

As with the question "How new is your theory?" the question "How nonobvious is your theory?" is best answered by our scholarly peers. If, when they are presented with your theory, they hit themselves in the head and say, "Wow, I never thought about it like that, but it makes a lot of sense!" then you have scored very well on this question.

Both of these last two questions illustrate an important part of the role of theory development in any science. It makes sense to think about theories as being like products and scientific fields as being very much like markets in which these products are bought and sold. Like other entrepreneurs in the marketplace, scientific entrepreneurs will succeed to the extent that their theories (products) are new and exciting (nonobvious). But, what makes a theory "new and exciting" is very much dependent on what has come before it.

2.9 CONCLUSION

We have presented a series of different strategies for developing theories of politics. Each of these strategies involves some type of thought exercise in which we arrange and rearrange our knowledge about the political world in hopes that doing so will lead to new causal theories. You have, we're certain, noticed that there is no simple formula for generating a new theory

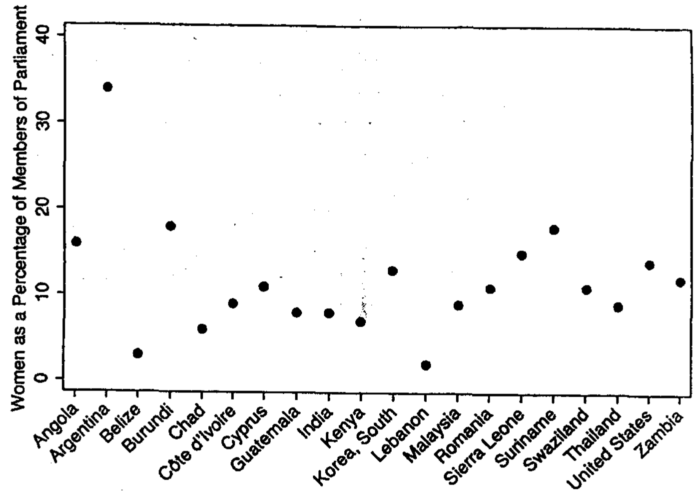

Figure 2.4. Women as a percentage of members of parliament, 2004.

and hopefully, as a result, appreciate our description of theory building as an "art" in the chapter's title. Theoretical developments come from many places and being critically immersed in the ongoing literature that studies your phenomenon of choice is a good place to start.

CONCEPTS INTRODUCED IN THIS CHAPTER

complete information
cross-sectional measure
expected utility
formal theory
incomplete information
intransitive
preference orderings
rational choice
rational utility maximizers
spatial dimension
strategic vote
time dimension
time-series measure
transitive
utility

EXERCISES

1. Figure 2.3 shows gross U.S. government debt as a percentage of GDP from 1960 to 2004. Can you think of a theory about what causes this variable to be higher or lower?
2. Figure 2.4 shows the percentage of a nation's members of parliament who were women for 20 randomly selected nations in 2004. Can you think of a theory about what causes this variable to be higher or lower?

3. Think about a political event with which you are familiar and follow these instructions:
   (a) Write a short description of the event.
   (b) What is your understanding of why this event happened the way that it happened?
   (c) *Moving from local to global*: Reformulate your answer to part (b) into a general causal theory without proper nouns.
4. Find a political science journal article of interest to you, and of which your instructor approves, answer the questions, and follow the instructions:
   (a) What is the main dependent variable in the article?
   (b) What is the main independent variable in the article?
   (c) Briefly describe the causal theory that connects the independent and dependent variables.
   (d) Can you think of another independent variable that is not mentioned in the article that might be causally related to the dependent variable? Briefly explain why that variable might be causally related to the dependent variable.
5. Imagine that the way in which the U.S. House of Representatives is elected was changed from the current single-member district system to a system of national proportional representation in which any party that obtained at least 3% of the vote nationally would get a proportionate share of the seats in the House. How many and what types of parties would you expect to see represented in the House of Representatives under this different electoral system? What theories of politics can you come up with from thinking about this hypothetical scenario?
6. *Applying formal theory to something in which you are interested.* Think about something in the political world that you would like to better understand. Try to think about the individual-level decisions that play a role in deciding the outcome of this phenomenon. What are the expected benefits and costs that the individual who is making this decision must weigh?

3 Evaluating Causal Relationships

OVERVIEW

Modern political science fundamentally revolves around establishing whether there are *causal relationships* between important concepts. This is rarely straightforward and serves as the basis for almost all scientific controversies. How do we know, for example, if economic development causes democratization, or if democratization causes economic development, or both, or neither? To speak more generally, if we wish to know whether some $X \rightarrow Y$, we need to cross four causal hurdles: (1) Is there a credible causal mechanism that connects X to Y? (2) Can we eliminate the possibility that Y causes X? (3) Is there covariation between X and Y? (4) Is there some Z related to both X and Y that makes the observed relationship between X and Y spurious? Many people, especially those in the media, make the mistake that crossing just the third causal hurdle – observing that X and Y covary – is tantamount to crossing all four. In short, finding a relationship is not the same as finding a *causal* relationship, and causality is what we care about as political scientists.

I would rather discover one causal law than be King of Persia.
– Democritus (quoted in Pearl 2000)

3.1 CAUSALITY AND EVERYDAY LANGUAGE

Like that of most sciences, the discipline of political science fundamentally revolves around evaluating causal claims. Our theories – which may be right or may be wrong – typically specify that some independent variable causes some dependent variable. We then endeavor to find appropriate empirical evidence to evaluate the degree to which this theory is or is not supported. But how do we go about evaluating causal claims? In this chapter and the next, we discuss some principles for doing this. We focus on the logic of causality and on several criteria for establishing with some confidence the degree to which a causal connection exists between two variables. Then, in Chapter 4, we discuss various ways to design research that help us to investigate causal claims. As we pursue answers to questions about causal relationships, keep our "rules of the road" from Chapter 1 in your mind, in particular the admonition to consider only empirical evidence along the way.

It is important to recognize a distinction between the nature of most scientific theories and the way the world seems to be ordered. Most of our theories are limited to descriptions of relationships between a *single* cause (the independent variable) and a *single* effect (the dependent variable). Such theories, in this sense, are very simplistic representations of reality, and necessarily so. In fact, as we noted at the end of Chapter 1, theories of this sort are laudable in one respect: They are parsimonious, the equivalent of bite-sized, digestible pieces of information. We cannot emphasize strongly enough that almost all of our theories about social and political phenomena are **bivariate** – that is, involving just two variables.

But social reality is *not* bivariate; it is **multivariate**, in the sense that any interesting dependent variable is caused by more than one factor. So although our theories describe the proposed relationship between some cause and some effect, we always have to keep in the forefront of our minds that the phenomenon we are trying to explain surely has many other possible causes. And when it comes time to design research to test our theoretical ideas – which is the topic of Chapter 4 – we have to try to account for, or "control for," those other causes. If we don't, then our causal inferences about whether our pet theory is right – whether X causes Y – may very well be wrong.[1] In this chapter we lay out some practical principles for demonstrating that, indeed, some X does cause Y. You also can apply these criteria when evaluating the causal claims made by others – be they a journalist, a candidate for office, a political scientist, a fellow classmate, a friend, or just about anyone else.

Nearly everyone, nearly every day, uses the language of causality – some of the time formally, but far more often in a very informal manner. Whenever we speak of how some event changes the course of subsequent events, we invoke causal reasoning. Even the word "because" implies that a causal process is in operation.[2] Yet, despite the ubiquitous

[1] Throughout this book, in the text as well as in the figures, we will use arrows as a shorthand for "causality." For example, the text "$X \rightarrow Y$" should be read as "X causes Y." Oftentimes, especially in figures, these arrows will have question marks over them, indicating that the existence of a causal connection between the concepts is uncertain.

[2] This example was suggested to us by Brady (2002).

use of the words "because," "affects," "impacts," "causes," and "causality," the meanings of these words are not exactly clear. Philosophers of science have long had vigorous debates over competing formulations of "causality."[3]

Although our goal here is not to wade too deeply into these debates, there is one feature of the discussions about causality that deserves brief mention. Most of the philosophy of science debates originate from the world of the physical sciences. The notions of causality that come to mind in these disciplines are mostly **deterministic** – that is, if some cause occurs, then the effect will occur *with certainty*. In contrast, though, the world of human interactions is **probabilistic** – increases in X are associated with increases (or decreases) in the probability of Y occurring, but those probabilities are not certainties. Whereas physical laws like Newton's laws of motion are deterministic – think of the law of gravity here – social science more closely resembles probabilistic causation like that in Darwin's theory of natural selection, in which random mutations make an organism more or less fit to survive and reproduce.[4] However, in reviewing three prominent attempts within the philosophy of science to elaborate on the probabilistic nature of causality, the philosopher Wesley Salmon (1993, p. 137) notes that "In the vast philosophical literature on causality [probabilistic notions of causality] are largely ignored." But in political science, our conceptions of causality must be probabilistic in nature. When we theorize, for example, that an individual's level of wealth causes her opinions on optimal tax policy, we do not at all mean that *every* wealthy person will want lower taxes, and *every* poor person will prefer higher taxes. Consider what *would happen if* we found a single rich person who favors high taxes or a single poor person who favors low taxes. One case alone does not decrease our confidence in the theory. In political science there will always be exceptions because human beings are not deterministic robots whose behaviors conform to lawlike statements. In other sciences in which the subjects of study are more robotic, it may make more sense to speak of laws that describe behavior. Consider the study of planetary orbits, in which scientists can precisely predict the movement of celestial bodies hundreds of years in advance. The political world, in contrast, is extremely difficult to predict. As a result, most of the time we are happy to be able to make probabilistic statements about causal relationships.

[3] You can find an excellent account of the vigor of these debates in a 2003 book by David Edmonds and John Eidinow titled *Wittgenstein's Poker: The Story of a Ten Minute Argument Between Two Great Philosophers*.

[4] We borrow the helpful comparison of probabilistic social science to Darwinian natural selection from Brady (2004).

What all of this boils down to is that the entire notion of what it means for something "to cause" something else is far from a settled matter. Should social scientists abandon all hope of finding causal connections? Not at all. What it means is that we should proceed cautiously and with an open mind, rather than in some hyperformulaic fashion.

3.2 FOUR HURDLES ALONG THE ROUTE TO ESTABLISHING CAUSAL RELATIONSHIPS

If we wish to investigate whether some independent variable, which we will call X, "causes" some dependent variable, which we will call Y, what procedures must we follow before we can express our degree of confidence that a causal relationship does or does not exist? Finding some sort of covariation (or, equivalently, correlation) between X and Y is not sufficient for such a conclusion.

We encourage you to bear in mind that establishing causal relationships between variables is not at all akin to hunting for DNA evidence like some episode from a television crime drama. Social reality does not lend itself to such simple, cut-and-dried answers. In light of the preceding discussion about the nature of causality itself, consider what follows to be guidelines as to what constitutes "best practice" in political science. With any theory about a causal relationship between X and Y, we should carefully consider the answers to the following four questions:

1. Is there a credible causal mechanism that connects X to Y?
2. Could Y cause X?
3. Is there covariation between X and Y?
4. Is there some **confounding variable** Z that is related to both X and Y and makes the observed association between X and Y spurious?

First, we must consider whether it is credible to claim that X *could* cause Y. To do this, we need to go through a thought exercise in which we evaluate the mechanics of how X would cause Y. In other words, what is it specifically about having more (or less) of X that will in all probability lead to more (or less) of Y? In effect, this hurdle represents an effort to answer the "how" and "why" questions about causal relationships. The more outlandish these mechanics would have to be, the less confident we are that our theory has cleared this first hurdle. Failure to clear this first hurdle is a very serious matter; the result being that either our theory needs to be thrown out altogether, or we need to revise it after some careful rethinking of the underlying mechanisms through which it works. It is worth proceeding to the second question only once we have a "yes" answer to this question.

Second, and perhaps with greater difficulty, we must ask whether it ossible (or even likely) that Y might cause X. As you will learn from : discussion of the various strategies for assessing causal connections in apter 4, this poses thorny problems for some forms of social science search, but is less problematic for others. Occasionally, this causal hurdle an be crossed logically. For example, when considering whether a person's gender (X) causes him or her to have particular attitudes about abortion policy (Y), it is a rock-solid certainty that the reverse-causal scenario can be dismissed: A person's attitudes about abortion does not "cause" them to be male or female. If our theory does not clear this particular hurdle, the race is not lost. Under these circumstances, we should proceed to the next question, while keeping in mind the possibility that our causal arrow might be reversed.

Throughout our consideration of the first two causal hurdles, we were concerned with only two variables, X and Y. The third causal hurdle can involve a third variable Z, and the fourth hurdle always does. Often it is the case that there are several Z variables.

For the third causal hurdle, we must consider whether X and Y covary (or, equivalently, whether they are correlated or associated). Generally speaking, for X to cause Y, there must be some form of measurable association between X and Y, such as "more of X is associated with more of Y," or "more of X is associated with less of Y." Demonstrating a simple bivariate connection between two variables is a straightforward matter, and we will cover it in Chapter 8. Of course, you may be familiar with the dictum "Correlation does not prove causality," and we wholeheartedly agree. It is worth noting, though, that correlation is normally an essential component of causality. But be careful. It is possible for a causal relationship to exist between X and Y even if there is no bivariate association between X and Y. Thus, even if we fail to clear this hurdle, we should not throw out our causal claim entirely. Instead, we should consider the possibility that there exists some confounding variable Z that we need to "control for" before we see a relationship between X and Y. Whether or not we find a bivariate relationship between X and Y, we should proceed to our fourth and final hurdle.

Fourth, in establishing causal connections between X and Y, we must face up to the reality that, as we noted at the outset of this chapter, we live in a world in which most of the interesting dependent variables are caused by more than one – often many more than one – independent variable. What problems does this pose for social science? It means that, when trying to establish whether a particular X causes a particular Y, we need to "control for" the effects of other causes of Y (and we call those other effects Z). If we fail to control for the effects of Z, we are quite likely to misunderstand the relationship between X and Y and make the wrong inference about whether X causes Y. This is the most serious mistake a social scientist can make. If we find that X and Y are correlated, but that, when we control for the effects of Z on both X and Y, the association between X and Y disappears, then the relationship between X and Y is said to be spurious.

3.2.1 Putting It All Together – Adding Up the Answers to Our Four Questions

As we have just seen, the process for evaluating a theoretical claim that X causes Y is a complicated process. Taken one at a time, each of the four questions in the introduction to this section can be difficult to answer with great clarity. But the challenge of evaluating a claim that X causes Y involves summing across all four of these questions to determine our overall confidence about whether X causes Y. To understand this, think about the analogy that we have been using by calling these questions "hurdles." In track events that feature hurdles, runners must do their best to try to clear each hurdle as they make their way toward the finish line. Occasionally even the most experienced hurdler will knock over a hurdle. Although this slows them down and diminishes their chances of winning the race, all is not lost. If we think about putting a theory through the four hurdles posed by the preceding questions, there is no doubt our confidence will be greatest when we are able to answer all four questions the right way ("yes," "no," "yes," "no") and without reservation. As we described in the introduction to this section, failure to clear the first hurdle should make us stop and rethink our theory. This is also the case if we find our relationship to be spurious. For the second and third hurdles, however, failure to clear them completely does not mean that we should discard the causal claim in question. Figure 3.1 provides a summary of this process. In the subsections that follow, we will go through the process described in Figure 3.1 with a series of examples.

3.2.2 Identifying Causal Claims Is an Essential Thinking Skill

We want to emphasize that the logic just presented does not apply merely to political science research examples. Whenever you see a story in the news, or hear a speech by a candidate for public office, or, yes, read a research article in a political science class, it is almost always the case that some form of causal claim is embedded in the story, speech, or article. Sometimes those causal claims are explicit – indented and italicized so that you just can't miss them. Quite often, though, they are harder to spot, and most of the time not because the speaker or writer is trying to confuse you.

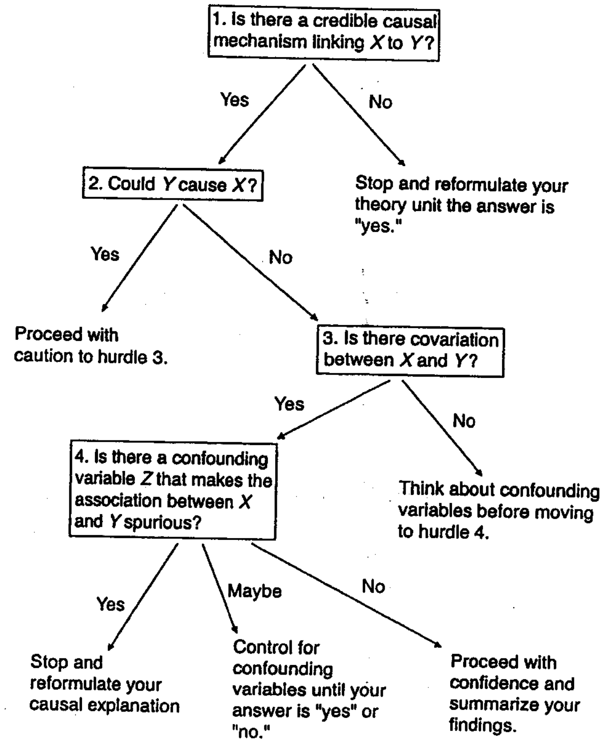

Figure 3.1. The path to evaluating a causal relationship.

What we want to emphasize is that spotting and identifying causal claims is a thinking skill. It does not come naturally to most people, but it can be practiced.

Take a common example from a political campaign: A candidate for president or prime minister who is running for reelection asserts that the voters should give him or her another term in office because the national economy is performing well. (Or, if economic performance is poor, the challenger will claim that the voters should replace the poor economic management team of the incumbent with the challenger's party.) There are perhaps two related causal claims embedded in a candidate's appeal to be reelected on the basis of the economy's performance. First, the candidate may be saying that the economic performance is better than it would be if the voters had chosen the other candidate in the last election. Second, the incumbent may be claiming that economic performance will be better in the future if he or she is reelected than it will be if the opposing candidate wins.

Because the second claim is about an unpredictable future, let's set it aside for the moment; it is interesting speculation, but there's no doubt that it is just speculation. Focus, then, on the first claim: that the economy is performing well because of the administration's economic policies. Is such a causal claim credible? For us to make such a judgment, we need to focus on our four causal hurdles. To start, we need to evaluate whether there is a credible causal mechanism that connects X (the administration and its policies) with Y (economic performance). Thinking through the mechanics of how this causal relationship would work is pretty straightforward. Presidents and prime ministers have a wide array of economic policy-making tools at their disposal. It seems pretty reasonable that the use of these tools could cause the economy to fare better or worse. *The answer to the first question is "yes."* Second, is it possible that Y causes X? This would mean that the current economy caused the administration's economic policies. In this case, we would need to figure out whether policy enactments did or did not precede good economic performance. Because there are so many different policies being enacted at various points in time, this would be a difficult question to answer. To be conservative, let's say that *the answer to the second question is "yes."* To answer our third question we need to figure out whether X (the administration and its policies) is associated with Y (economic performance). Presumably it is, and presumably this relationship is such that the economic performance improved after the administration's policies were put in place – right? – or else the candidate would not be making the claim. But, if we wanted to evaluate such a claim, we would have to choose an indicator of economic performance and make a comparison across time. For now, let's say that *the answer the third question is "yes."* So far, the politician's claim of a causal relationship is doing pretty well (because our answers are "yes," "yes," and "yes"). But the fourth causal hurdle is where the candidate (and we) can get tripped up. Is there some other force, Z, that is related to both X and Y and renders the relationship between X and Y spurious? Think about this one: Can we think of any other reasons, besides administration policies, why the economy might be performing well? Of course we can. It would be patently silly to assert that the *sole cause* of strong economic performance is government policy. Innovation in the private sector (Z), for example, might (by happenstance) coincide with government policy changes (X) and be strongly related to prosperity (Y). Without some further analysis, we have ample reason to be skeptical of such a candidate's claim that the economy is prosperous because of the administration's sound policies. Unfortunately for the politician, *the answer to the fourth question is "maybe."* We're going to have to see some more evidence before we, as scientists, are going to believe her causal claim.

We could rather easily think of a host of other factors that might fit the description of a confounding variable. But let's be careful before we dismiss the politician as a charlatan. Does this mean that the candidate is wrong, and that we know that administration policies did *not* cause prosperity? Absolutely not. All we have done in this simple thinking exercise is to recognize that the dependent variable of interest (Y), the health of the economy, is certainly a function of many things, one of which may or may not be the administration's economic policies (X). To know that the administration produced the prosperity, we would need to *control for* other possible causes of prosperity, and we haven't done that. Therefore we should conclude that it is possible that the candidate's claim is appropriate. But it has not yet been empirically supported, because alternative explanations have not yet been ruled out. Identifying the underlying causal claim, in this case, helps us to be skeptical of the self-interested claims of political actors. A candidate's job, of course, is not to evaluate causal claims carefully; it is to get votes. But evaluating the credibility of a candidate's often-implicit causal claims is important if we, the voters, do not want to be led astray by vote-hungry politicians.

An important part of taking a scientific approach to the study of politics is that we turn the same skeptical logic loose on scholarly claims about causal relationships. Before we can evaluate a causal theory, we need to consider how well the available evidence answers each of the four questions about X, Y, and Z. Once we have answered each of these four questions, one at a time, we then think about the overall level of confidence that we have in the claim that X causes Y.

3.2.3 What Are the Consequences of Failing to Control for Other Possible Causes?

When it comes to any causal claim, as we have just noted, the fourth causal hurdle often trips us up, and not just for evaluating political rhetoric or stories in the news media. This is true for scrutinizing scientific research as well. In fact, a substantial portion of disagreements between scholars boils down to this fourth causal hurdle. When one scholar is evaluating another's work, perhaps the most frequent objection is that the researcher "failed to control for" some potentially important cause of the dependent variable.

What happens when we fail to control for some plausible other cause of our dependent variable of interest? Quite simply, it means that we have failed to cross our fourth causal hurdle. *So long as a credible case can be made that some uncontrolled-for Z might be related to both X and Y, we cannot conclude with full confidence that X indeed causes Y.* Because the main goal of science is to establish whether causal connections between variables exist, then failing to control for other causes of Y is a potentially serious problem.

One of the themes of this book is that statistical analysis should not be disconnected from issues of research design – such as controlling for as many causes of the dependent variable as possible. When we discuss multiple regression (in Chapters 10 and 11), which is the most common statistical technique that political scientists use in their research, the entire point of those chapters is to learn how to control for other possible causes of the dependent variable. We will see that failures of research design, such as failing to control for all relevant causes of the dependent variable, have statistical implications, and the implications are always bad. Failures of research design produce problems for statistical analysis, but hold this thought. What is important to realize for now is that good research design will make statistical analysis more credible, whereas poor research design will make it harder for any statistical analysis to be conclusive about causal connections.

3.3 WHY IS STUDYING CAUSALITY SO IMPORTANT? THREE EXAMPLES FROM POLITICAL SCIENCE

Our emphasis on causal connections should be clear. We turn now to several active controversies within the discipline of political science, showing how debates about causality lie at the heart of precisely the kinds of controversies that got you (and most of us) interested in politics in the first place.

3.3.1 Life Satisfaction and Democratic Stability

One of the enduring controversies in political science is the relationship between *life satisfaction in the mass public* and *the stability of democratic institutions*. Life satisfaction, of course, can mean many different things, but for the current discussion let us consider it as varying along a continuum, from the public's being highly unsatisfied with day-to-day life to being highly satisfied. What, if anything, is the causal connection between the two concepts?

Political scientist Ronald Inglehart (1988) argues that life satisfaction (X) *causes* democratic system stability (Y). If we think through the first of the four questions for establishing causal relationships, we can see that there is a credible causal mechanism that connects X to Y – if people in a democratic nation are more satisfied with their lives, they will be less likely to want to overthrow their government. *The answer to our first question is "yes."* Moving on to our second question: Is it possible that democratic stability (Y) is what causes life satisfaction (X)? Certainly it is. It is very easy

to conceive of a causal mechanism in which citizens take careful note of the political system when they consider how happy they are and that citizens living in stable democracies are apt to look back on a history of government stability – that is, a recent history without violent revolutions – and feel a sense of safety and happiness as a result. *The answer to our second question is "yes."* We now turn to the third question. Using an impressive amount of data from a wide variety of developed democracies, Inglehart and his colleagues have shown that there is, indeed, an association between average life satisfaction in the public and the length of uninterrupted democratic governance. That is, countries with higher average levels of life satisfaction have enjoyed longer uninterrupted periods of democratic stability. Conversely, countries with lower levels of life satisfaction have had shorter periods of democratic stability and more revolutionary upheaval. *The answer to our third question is "yes."* With respect to the fourth question, it is easy to imagine a myriad of other factors (Z's) that lead to democratic stability, and whether Inglehart has done an adequate job of controlling for those other factors is the subject of considerable scholarly debate. *The answer to our fourth question is "maybe."* Inglehart's theory has satisfactorily answered questions 1 and 3, but it is the answers to questions 2 and 4 that have given skeptics substantial reasons to doubt his causal claim.

3.3.2 School Choice and Student Achievement

In recent years, during which there has been considerable concern about the performance of public elementary and secondary schools, the possibility of the government issuing vouchers to allow families to send children to private schools has become highly controversial. Setting the normative issues aside about whether "school choice" is either inherently desirable or instead something that will by its nature drain the public schools, there is lurking in the background an important empirical and causal issue: Does the type of school a child attends (X) affect student performance (Y)? It can be argued that, as researchers cannot demonstrate that school-choice programs improve student performance, the programs lose a substantial portion of their appeal.

Clearly, the first question establishing causal relationships is easy enough to answer, because a credible (if not airtight) argument can be made that children will receive an education that better prepares them for standardized tests in private schools, which typically have smaller class sizes and fewer layers of bureaucracy. *The answer to our first question is "yes."* In this example, the second hurdle is pretty easy to clear – how could test-score results (Y) *cause* the type of school (X)? *The answer to our second question is "no."*

Let's move to the third question of whether there is covariation between X and Y. At first glance, this would seem like an entirely straightforward matter. Find a city or state where there is a school-choice program; compare the scores on standardized tests between students in the public school with those in the private school; then draw a conclusion. Is this comparison useful? Suppose we were to compare scores on a standardized math test among eighth-graders in Metropolis, USA, some of whom went to private schools by way of a school-choice program and others who remained in Metropolis's public schools. And suppose we find that, indeed, the average math test score among students who participate in the choice program is higher than that of those who remained in the public school. In this hypothetical case, *the answer to our third question is "yes."* Our theory is looking pretty good. So far, all of the answers have supported it. Does this mean that the choice program *caused* their test scores to be higher?

It is a tempting conclusion to leap to, isn't it? It sounds like a classic case of comparing apples to apples, so to speak. But let's try to stick to our four questions. We have already, in our hypothetical example, conceded that the type of school (X) is associated with test scores (Y).

The fourth question is the only one remaining, and it is, in this case, a difficult question to answer. Can you think of another cause (Z) that is related to whether or not a student enrolls in the choice program (X) that will also be related to the standardized test score? Yes. In this case, the level of parental involvement (Z) could surely affect both X and Y and might make the association that we see between X and Y spurious. Parents who are actively involved in their child's education (Z) are more likely to be aware of a school-choice program in their district and are more likely to pursue that option (X). Similarly, parents with high levels of involvement in their child's education (Z) are more likely to have children who perform well on standardized tests (Y); such parents read to their children more, help them with homework, and stress the importance of education in a child's life.

In this case, the Z variable we have identified produces what is called a **selection effect** – a situation in which a systematic force causes only a nonrandom subset of eligible targets to participate in a program. In any substantive area in which we are trying to evaluate the effectiveness of a government policy, it is critical to compare participants in the program with nonparticipants in a rigorous fashion. If we find systematic differences between participants and nonparticipants – as we surely would in a school-choice program – then it becomes *exceedingly crucial* to try to control for those forces when evaluating the program's effectiveness. In the school-choice example considered here, what seemed like a simple apples-to-apples comparison really turned out to be an apples-to-giraffes comparison. At the very least, *the answer to our first question is "maybe."*[5]

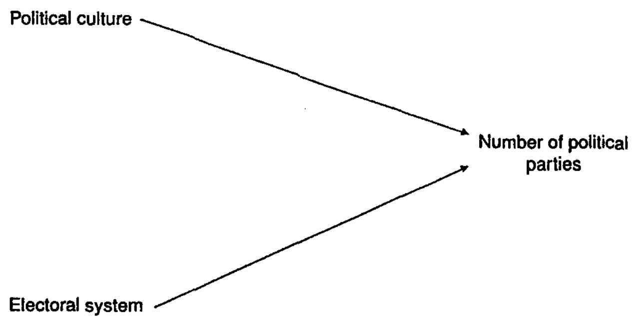

Figure 3.2. Theoretical causes of the number of parties in legislatures.

Let's be extremely careful here. Does this mean that school-choice programs do not help to improve student performance on tests? Not at all. What our four questions do is remind us that, sometimes, the temptingly easy conclusion needs additional scrutiny before we embrace it. In Chapter 4, we will talk about some research designs that can help to ameliorate situations precisely like this one.

3.3.3 Electoral Systems and the Number of Political Parties

Political science has a long tradition of examining the impact of institutional arrangements on political outcomes. One prominent example of this type of research has focused on the influence of electoral systems on the number of political parties in legislatures. Figure 3.2 depicts a theoretical model of the number of parties that will be represented in a legislature. The first theory is that, the more societal divisions there are that shape a political culture, the more political parties there will be in the legislature. The second theory on which we focus in this subsection is that, if we hold constant the political culture of the area that the legislature represents, the more disproportional the electoral system is in translating votes into seats (X) and fewer political parties will be represented in the legislature (Y).

The term "disproportional" in this theory is expressed in terms of the translation of votes into seats for political parties. A perfectly proportional system would be one in which the percentage of votes cast for each party was *exactly* equal to the percentage of seats awarded to that party in the

[5] In addition to parental involvement, there are other possible selection mechanisms at work. A private school involved in a school-choice program, for example, might choose to use test scores as a criterion for admission.

legislature as a result of the election. In practice, perfectly proportional electoral systems are never found; in fact electoral systems differ substantially in terms of how close they come to this ideal. Turning to our four hurdles, the causal mechanism behind the theory of electoral systems and the number of parties is driven by the organizational incentives politicians face when deciding whether to form new political parties or work within established parties to contest elections. Disproportionate electoral systems tend to reward the largest parties and greatly penalize the smaller parties in terms of translating votes into seats. Thus, the more disproportionate the electoral system, the greater will be the tendency for politicians who are competing for legislative seats to band together, resulting in fewer political parties in the legislature. If you believe this, then *the answer to our first question is "yes."*

To better understand this theory of the influence of electoral institutions, consider the U.S. House of Representatives. A quick review of the history of party membership in the U.S. House indicates that, with few exceptions, two political parties have held all or most of the seats. According to this theory, this is the case *because* the U.S. House of Representatives is elected by use of a set of rules that produce disproportionate outcomes in terms of the translation of votes to seats. That system is known as a "single-member district plurality" system. The entire country is divided into electoral districts, and on election day whichever candidate receives the most votes (a plurality) is elected to represent that district. When the votes and seats are tallied up at the national level, results tend disproportionately to favor the parties with the most votes. For example, in the 1992 U.S. House elections, Democratic Party candidates received 49.95% of the votes cast and 59.31% of the seats. In that same election, Republican Party candidates received 44.75% of the votes cast and 40.46% of the seats; all other parties together received 5.3% of the votes cast and only one seat (or 0.2% of the available seats).

One of the most proportional electoral systems in history was that of the Weimar Republic in Germany between World War I and World War II. Under the Weimar Republic electoral system, Germany was divided into 13 electoral regions, and seats in the national legislature were awarded to any party that managed to get 60,000 or more votes in any one of the electoral regions. Given that the number of voters who turned out in Weimar Republic elections was never less than 28 million, politicians had very little legal incentive to band together to contest elections. Consistent with the theory, the Reichstag had many different political parties throughout the time of the Weimar Republic. Some scholars have suggested that, because the politicians were divided into so many different political parties in the legislature, they were unable to band together to counter the

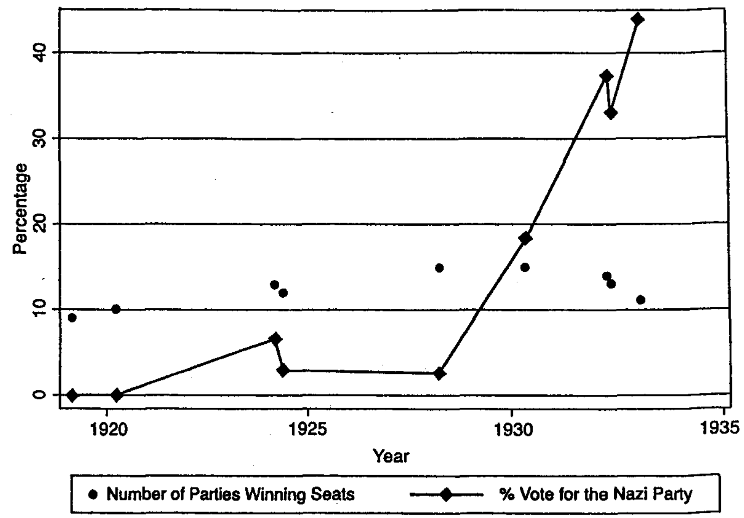

Figure 3.3. Nazi vote and the number of parties winning seats in Weimar Republic elections, 1919–1933.

rising strength and popularity of the Nazi Party. The Nazi Party, in contrast with many of the other parties in Germany at the time, was willing and able to hold together its politicians through coercive means. Figure 3.3 shows the number of parties in the Reichstag and the percentage of votes for the Nazi Party across the period of the Weimar Republic.

In the aftermath of World War II, the constitution of West Germany was designed with the Weimar Republic experience very much in mind. Although the electoral system continued to be a form of proportional representation, one major change was that parties that got less than 5% of the vote nationally were not given seats in the national legislature.[6] If we look at Figure 3.4, we can see an interesting pattern. In the first election after World War II, 11 political parties won seats in the Bundestag. After this, though, fewer and fewer political parties were represented, with only four parties in the Bundestag throughout the 1960s and 1970s. In the 1980s, the political culture of what was then West Germany began to change. The Green Party cleared the 5% threshold and was represented in the Bundesrat. In 1990 East and West Germany were reunified. In each of the elections since then,

[6] The West German electoral system, which is now the electoral system for reunified Germany, is one of the most complicated electoral systems in the world. It is possible for a party to gain seats despite failing to clear the 5% threshold provided that they win district-level seats. For a nice overview of this system, we recommend Gallagher, Laver, and Mair (2006).

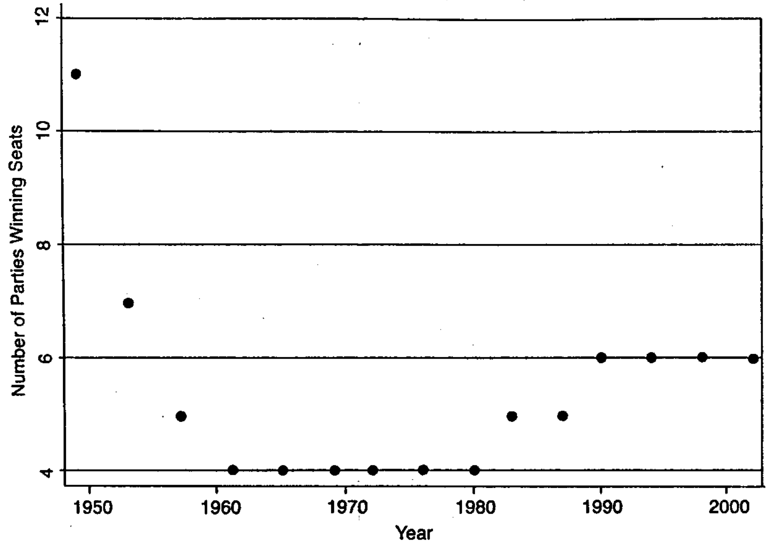

Figure 3.4. Number of parties winning seats in German Bundestag elections, 1949–2002.

six parties have been represented in the Bundesrat. The additional party is the Party of Democratic Socialism, which is a left-wing party created out of the remains of the East German Communist Party.

Consider the evidence from Germany in terms of our four questions. In the preceding subsections, we have a reasonably credible mechanism (incentives faced by office-seeking politicians) of how X (type of electoral system) causes Y (the number of political parties), and thus we answer our first question. We do not have any evidence from this case that the number of political parties *caused* the electoral system so we can be confident that *the answer to our second question is "no."* Figures 3.3 and 3.4 certainly seem to indicate that there was covariation between the electoral system and the number of parties.[7] When the more disproportionate electoral system of post–World War II Germany was put into effect, the number of political parties went down substantially. On the basis of this evidence we can conclude preliminarily that *the answer to our third question is "yes."* We have not yet conducted an extensive search for confounding variables (Z) that may be related to both the electoral system (X) and the number of political parties (Y). But it is difficult to imagine such a variable. So, on the basis of our consideration of our theories so far, *the answer to our fourth question is "no."* Taken together, this theory has done very well as

[7] Beginning in Chapter 8 we will discuss more systematic ways in which to use statistical techniques to evaluate empirical evidence of relationships between variables.

we have subjected it to the four causal hurdles. But, as we will learn later, we should base our answer to question 3 on more evidence than what we have examined thus far.

3.4 WHY IS STUDYING CAUSALITY SO IMPORTANT? THREE EXAMPLES FROM EVERYDAY LIFE

Causal claims are not limited to social science research like those previously discussed. There are times when causal claims in politics, the news, or just everyday life are downright humorous. Learning the intellectual habit of sifting through an argument to find the embedded causal argument can be useful.

3.4.1 Alcohol Consumption and Income

When you're in the checkout line at the grocery store, do you ever pick up the tabloids and scan them for the latest news about alien abductions and celebrity breakups? If so, you might remember this gem in the tabloid magazine *Weekly World News* of May 14, 2002, under the screaming headline "Want to be loaded? Then get loaded!":

> Do you want to be rich and successful, like Donald Trump, Bill Gates or Oprah Winfrey? Then belly up to the bar and drink your way to wealth!... when it comes to raking in the dough, boozers leave teetotalers in the dust, with all but the heaviest drinkers earning more.

Perhaps it's easy to believe that such a claim would appear in a tabloid magazine. At least, in this case, the causal claim is right there in the title of the article. Think about it for a moment: Given the fact that the consumers of tabloids do not have a bevy of data at their disposal about both alcohol consumption and adult earnings, what kind of evaluation can we make about such a causal claim? Think about our causal hurdles. Perhaps – perhaps! – we can cross the third hurdle by finding that it is true that alcohol consumption (X) is associated with higher earnings (Y). What about our second hurdle? Is it possible that earnings cause alcohol consumption? It is, at least in the sense that individuals with higher incomes have more discretionary dollars to spend on anything – including alcohol – that they like, and that, in contrast, people with lower levels of income have a natural ceiling on how much they can spend on alcohol. So maybe the causal arrow runs the other way after all, though not in the pernicious way that the tabloid suggests. The fourth causal hurdle – trying to think of possible confounding variables that might be related to both X and Y – is also simple in this case: People who work in corporate America and have business dinners with clients are more likely to consume alcohol, and they are also apt to make higher salaries than those who are not in the corporate world. But, most egregiously, the first causal hurdle trips us up. As much as avid producers and consumers of alcoholic beverages might like to convince people that higher levels of drinking will translate to higher income, can you think of a *credible* causal mechanism that connects alcohol consumption to income?

3.4.2 Treatment Choice and Breast Cancer Survival

In 2006 the National Breast Cancer Foundation forecast that 211,000 women and 1600 men in the United States would be told by their doctors that they have breast cancer. One of the most painful situations that some patients have to face occurs when they have to choose a treatment option to pursue.

Two treatment strategies strike different balances between the desirability of aggressively treating the cancer and keeping the procedures as minimally invasive as possible. The first strategy, called a radical mastectomy, represents an effort to try to purge the entire body of cancer by removing the entire breast. In effect, this strategy acknowledges that the breast can produce cancer, so the choice is to remove it. Because of this, it is maximally invasive to a patient's body, and, understandably, few patients find it appealing. The second strategy, called a lumpectomy, is a more localized surgery in which the cancerous tumor is removed from the patient's breast, but as much of the breast as possible is left intact. Certainly this treatment is less invasive, less aggressive, and is less distasteful to most breast cancer patients. It also carries with it the risk that some cancer cells might be missed during the surgery and, as a result, left behind in the body.

Which treatment option should a patient choose? Obviously, patients might consider a myriad of factors when faced with such a choice, but among them might be the expected survival rates for each treatment option. They might expect that patients who choose radical mastectomies, on average, live longer than patients who choose lumpectomies, for the simple fact that the lumpectomy procedure carries with it the risk that some cancer cells will be missed and left behind in the breast tissue, a risk that the radical mastectomy, by its very nature, avoids.

Perplexingly, there tends to be *no association* between breast cancer treatment choice (X) and posttreatment longevity (Y). That is, patients who go through both procedures have roughly the same 1-year and 5-year survival rates. (This is to say, the third causal hurdle has not been crossed.) Does this mean that radical mastectomies are unnecessarily invasive and should not be considered a good treatment option?

Again, consider the rest of our causal hurdles. From the preceding discussion of treatments, it is pretty clear that we can clear the first hurdle – there is a credible causal mechanism between X and Y. The second hurdle, that longevity (Y) might cause treatment (X), is obviously not possible. But the fourth causal hurdle is crucial for evaluating this relationship. Can we imagine any factors (Z) that might be related to both treatment choice (X) and longevity (Y)? Certainly we can. Not all cancers are detected at the same stage of advancement. Some are caught early, whereas others are detected only when the cancer has spread considerably. Thus the severity of cancer at detection (Z) might affect both treatment choice and longevity. Patients whose cancers are diagnosed in early stages (Z), when the tumors are small, may be more likely to choose lumpectomies (X) and are more likely to survive (Y). Conversely, patients whose cancer is spotted in advanced stages (Z) may have almost no choice but to get the more radical treatment (X), and their prospects for long-term survival (Y) are less bright.

Like the school-choice example discussed earlier, this is a case in which a third variable – severity of the disease when it is detected – operates as a selection mechanism that makes the comparison between individuals with different values of the independent variable – treatment – extremely difficult to compare. Although it might be tempting simply to examine the bivariate relationship between treatment and survival rates and conclude that the treatments don't produce different results, that conclusion might be exactly wrong. Why? Because the patients who get radical mastectomies are systematically different from those who get lumpectomies. Simple comparisons in such a case can produce incorrect inferences about causal effects.[8]

Note that this is one of those somewhat unusual situations in which we believe that X may indeed cause Y ***in spite of the fact that we did not successfully clear the third causal hurdle*** – that is, that there is no bivariate association between X and Y. This supports our view that once a theory has successfully cleared the first hurdle (meaning that there is a credible causal mechanism) it should be put through all of the remaining three causal hurdles.

3.4.3 Explicit Lyrics and Teen Sexual Behavior

What is the role of popular culture in determining teen behavior? Is it the case that the explicit sexual content that saturates so much of today's

[8] It should be obvious that we are not oncologists with particular expertise in this very sensitive area. We are not in any way advocating or disparaging a particular treatment for breast cancer. Rather, we think it's important to show how thinking rigorously about causality can lead us to look past surface relationships and dig deeper to find better answers.

culture causes teenagers (especially) to be sexually active at an earlier age? Or is it the case that popular culture is merely a mirror, reflecting back to us who we truly are? A study reported by the Associated Press in 2006, titled "Sexual lyrics prompt teens to have sex," takes a rather clear position on this issue:

> Teens whose iPods are full of music with raunchy, sexual lyrics start having sex sooner than those who prefer other songs, a study found.... Teens who said they listened to lots of music with degrading sexual messages were almost twice as likely to start having intercourse or other sexual activities within the following two years as were teens who listened to little or no sexually degrading music. Among heavy listeners, 51 percent started having sex within two years, versus 29 percent of those who said they listened to little or no sexually degrading music.

So the third causal hurdle – whether *X* (music listening) and *Y* (sexual behavior) are related – has been been cleared. And, for the moment, let's dismiss the reverse-causal scenario (question 2) that a teen's sexual behavior causes them to listen to particular kinds of music.

But focus on the fourth causal question. Surely explicit lyrics cannot be the sole factor that causes teens to be sexually active. (And it's worth noting that no person in the article makes the claim that it is the only cause.) Are there other factors that might be related to both music-listening habits and sexual behaviors? According to the article, the research "tried to account for other factors that could affect teens' sexual behavior, including parental permissiveness, and still found explicit lyrics had a strong influence." Surely, parental permissiveness (*Z*) could be related to both music listening (*X*) and sexual behavior (*Y*), and the finding that the *X*–*Y* connection survived such a control is helpful. But are there *still other* possible causes in addition to parental permissiveness? Certainly, and critics mentioned in the article are quick to point them out. Could peer pressure (*Z*) be related to both *X* and *Y*? Absolutely. What about self-esteem? Again, yes. Failing to account for those possible causes – and any others that you might think of that can be related to both *X* and *Y* – can cause us to make the wrong inference about whether exposure to lyrics causes sexual behavior.

With respect to the first question – the existence of a credible causal mechanism – the article quotes a psychologist who sees a logical connection:

> The brain's impulse-control center undergoes "major construction" during the teen years at the same time that an interest in sex starts to blossom.... Add sexually arousing lyrics and "it's not that surprising that a kid with a heavier diet of that... would be at greater risk for sexual behavior."

Other psychologists might disagree, of course. But the failure to control for all other confounding variables that might be related to the independent and dependent variables is more than enough ammunition to allow a savvy record-company executive to cast doubts on such a study.

3.5 WRAPPING UP

Learning the thinking skills required to evaluate causal claims as conclusively as possible requires practice. They are intellectual habits that, like a good knife, will sharpen with use.

Translating these thinking skills into actively designing new research that helps to address causal questions is the subject of Chapter 4. All of the "research designs" that you will learn in that chapter are strongly linked to issues of evaluating causal claims. Keeping the lessons of this chapter in mind as we move forward is essential to making you a better consumer of information, as well as edging you forward toward being a producer of research.

CONCEPTS INTRODUCED IN THIS CHAPTER

bivariate	probabilistic
confounding variable	selection effect
deterministic	spurious
multivariate	

EXERCISES

1. Think back to a history class in which you learned about the "causes" of a particular historical event (for instance, the Great Depression, the French Revolution, or World War I). How well does each causal claim perform when you try to answer the four questions for establishing causal relationships?
2. Go to your local newspaper's web site (if it has one; if not, pick the web site of any media outlet you visit frequently). In the site's "Search" box, type the words "research cause" (without quotes). (*Hint:* You may need to limit the search time frame, depending on the site you visit.) From the search results, find two articles that make claims about causal relationships. Print them out, and include a brief synopsis of the causal claim embedded in the article.
3. For each of the following examples, imagine that some researcher has found the reported pattern of covariation between *X* and *Y*. Can you think of a variable *Z* that might make the relationship between *X* and *Y* spurious?
   (a) The more firefighters (*X*) that go to a house fire, the greater property damage that occurs (*Y*).

(b) The more money spent by an incumbent member of Congress's campaign (X), the lower their percentage of vote (Y).
(c) The more children in a community that participate in a Head Start program (X), the greater percentage of students that demonstrate kindergarten readiness (Y).
(d) The higher the salaries of Presbyterian ministers (X), the higher the price of rum in Havana (Y).

4. For each of the following pairs of independent and dependent variables, write about both a probabilistic and a deterministic relationship to describe the likely relationship:

(a) A person's education (X) and voter turnout (Y).
(b) A nation's economic health (X) and political revolution (Y).
(c) Candidate height (X) and election outcome (Y).

5. Take a look at the codebook for the data set "BES 2005 Subset" and write about your answers to the following items:

(a) Develop a causal theory about the relationship between an independent variable (X) and a dependent variable (Y) from this data set. Is it the credible causal mechanism that connects X to Y? Explain your answer.
(b) Could Y cause X? Explain your answer.
(c) What other variables (Z) would you like to control for in your tests of this theory?

4 Research Design

OVERVIEW

Given our focus on causality, what research strategies do political scientists use to investigate causal relationships? Generally speaking, the controlled experiment is the foundation for scientific research. And some political scientists use experiments in their work. However, owing to the nature of our subject matter, most political scientists adopt one of two types of "observational" research designs that are intended to mimic experiments. The cross-sectional observational study focuses on variation across individual units (like people or countries). The time-series observational study focuses on variation in aggregate quantities (like presidential popularity) over time. What is an "experiment" and why is it so useful? How do observational studies try to mimic experimental designs? Most important, what are the strengths and weaknesses of each of these three research designs in establishing causal relationships between concepts? That is, how does each one help us to get across the four causal hurdles identified in Chapter 3? Relatedly, we introduce issues relating to the selection of samples of cases to study in which we are not able to study the entire population of cases to which our theory applies. This is a subject that will feature prominently in many of the subsequent chapters.

4.1 COMPARISON AS THE KEY TO ESTABLISHING CAUSAL RELATIONSHIPS

So far, you have learned that political scientists care about causal relationships. You have learned that most phenomena we are interested in explaining have multiple causes, but our theories typically deal with only one of them while ignoring the others. In some of the research examples in the previous chapters, we have noted that the multivariate nature of the world can make our first glances misleading. In the breast cancer example, at first it did not appear that any kind of relationship (let alone a causal relationship) existed between treatment choice and patient longevity. In the school-choice example, it first appeared that a relationship (and perhaps a causal one) did exist between participation in the program and test scores. But, we argued, in both cases those first glances were potentially quite misleading.

Why? Because what appeared to be the straightforward comparisons between two groups – patients who chose one treatment compared with patients who chose another, or eighth-graders in one school compared with eighth-graders in another school – ended up being far from simple. On some very important factors, our different groupings for our independent variable X were far from equal. That is, patients who chose different treatment options (X) had differing levels of the disease when it was discovered (Z), which also affected their longevity (Y). And students in different school programs (X) had parents who had systematically different levels of involvement in their childrens' education (Z), which also affected test scores (Y). As convincing as those bivariate comparisons might have been, they would likely be misleading.

Comparisons are at the heart of science. If we are evaluating a theory about the relationship between some X and some Y, the scientist's job is to do everything possible to make sure that no other influences (Z) interfere with the comparisons that we will rely on to make our inferences about a possible causal relationship between X and Y.

The obstacles to causal inference that we described in Chapter 3 are substantial, but surmountable. We don't know whether, in reality, X causes Y. We may be armed with a theory that suggests that X does, indeed, cause Y, but theories can be (and often are) wrong or incomplete. So how do scientists generally, and political scientists in particular, go about testing whether X causes Y? There are several strategies, or research designs, that researchers can use toward that end. The goal of all types of research designs are to help us evaluate how well a theory fares as it makes its way over the four causal hurdles – that is, to answer as conclusively as is possible the question about whether X causes Y. In the next two sections we focus on the two strategies that political scientists use most commonly and effectively: experiments and observational studies.

4.2 EXPERIMENTAL RESEARCH DESIGNS

Suppose that you were the CEO of a pharmaceutical company, and your scientific team tells you that they have just discovered a new drug that will help lower blood pressure. The pharmacists tell you that they have

successfully tested the drug on rats and developed a dosage regimen that they expect will be effective on people. However, the drug has yet to be tested on people.

And it is important to add here that the causal claim has a particular directional component to it; that is, increased (not decreased) amounts of the drug are alleged to lower (not raise) blood pressure.

How would researchers in the physical sciences and medicine evaluate whether this new and promising drug works on humans? Note the focus on causality here. In more "causal" language, how can we find out whether taking the drug (X) will *cause* patients to have lower blood pressure (Y)? As the introduction to this chapter highlights, we will need a comparison of some kind, and we will want that comparison to isolate any potentially different effects that the drug has on a patient's blood pressure. It is very important, and not at all surprising, to realize that patients may have high or low blood pressure for a variety of reasons (Z's) that have nothing to do with our new drug – varying exercise habits, varying diets, and varying genetic predispositions can all cause blood pressure to be high or low. So how can we establish whether, among these other influences (Z), our new drug (X) also causes a patient's blood pressure (Y) to fall?

The standard answer to this question in the physical and medical sciences is that we would need to conduct an **experiment**. Because the word "experiment" has such common usage, its scientific meaning is frequently misunderstood. An experiment is *not* simply any kind of analysis that is quantitative in nature; neither is it exclusively the domain of laboratories and white-coated scientists with pocket protectors. We define an experiment as follows: An experiment is a research design in which the researcher both controls and randomly assigns values of the independent variable to the participants.

Notice the twin components of the definition of the experiment: That the researcher both *controls* values of the independent variable – or X, as we have called it – as well as *randomly assigns* those values to the participants in the experiment. Together, these two features form a complete definition of an experiment, which means that there are no other essential features of an experiment beside these two.

What does it mean to say that a researcher "controls" the value of the independent variable that the participants receive? It means, most important, that the values of the independent variable that the participants receive are *not* determined either by the participants themselves or by nature. In our example of our blood-pressure drug, this requirement means that we cannot compare people who, by their own choice, already take the drug with those who do not (in this case the choice of whether or not to take the drug is a Z variable that may exert an influence on Y separate from X). It means that we, the researchers, have to decide which of our experimental participants will take the drug and which ones will not.

But the definition of an experiment has one other essential component as well: We, the researchers, must not only control the values of the independent variable, but *we must also assign those values to participants randomly*. In the context of our drug-testing example, this means that we must toss coins, draw numbers out of a hat, use a random-number generator, or some other such mechanism to ensure that our participants are divided into a **treatment group** (who will receive our drug) and a **control group** (who will not receive the drug, but will instead presumably receive a placebo).

What's the big deal here? Why is randomly assigning subjects to treatment groups important? What scientific benefits arise from the random assignment of people to treatment groups? To see why this is so crucial, recall that we have emphasized that all science is about comparisons and also that every interesting phenomenon worth exploring – every interesting dependent variable – is caused by many factors, not just one. Random assignment to treatment groups ensures that the comparison we make between the treatment group and the control group is as pure as possible and that some other cause of the dependent variable (Z) will not pollute that comparison. By first taking a group of participants and then randomly splitting them into two groups on the basis of a coin flip, what we have ensured is that the participants will not be systematically different from one another. Indeed, in the aggregate – and provided that the participant pool is reasonably large – randomly assigning participants to treatment groups ensures that the groups, as a whole, are *identical*. If the two groups are identical, save for the coin flip, then we can be certain that any differences we observe in the groups must be because of the independent variable that we have assigned to them.

Return to our drug-trial example. An experiment involving our new blood-pressure drug would involve finding a group of people – however obtained – and then randomly assigning them to receive either the new drug or a placebo. We fully realize that there are other causes of low and high blood pressure and that our experiment does not negate those factors. In fact, our experiment will have nothing whatsoever to say about those other causes. What it *will* do, and do well, is to determine whether our drug has an effect on blood pressure.

Contrast the comparison that results from an experiment with a comparison that arises from a nonexperiment. (Be patient. We'll talk all about nonexperimental designs in the next section.) Suppose that the makers of a particular brand of aspirin wanted to test the claim that people who take their aspirin have lower blood pressure than people who don't take their

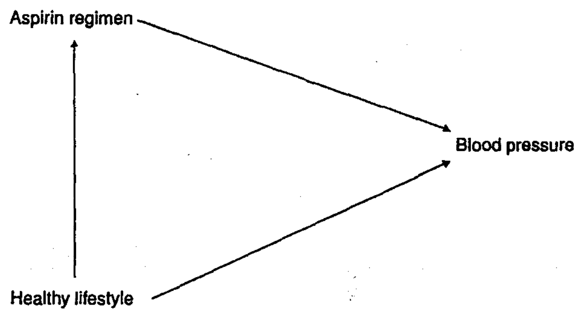

Figure 4.1. The possibly confounding effects of a healthy lifestyle on the aspirin–blood-pressure relationship.

aspirin. Let's even assume that they conduct a random-sample survey of adults, and that people answer the survey truthfully about their aspirin intake and blood pressure. If there is an elevated rate of high blood pressure in the nonaspirin group compared with that of the aspirin takers, does that mean that aspirin *caused* – see that word again – people to have lower blood pressure? No.[1] Why not? Because aspirin-takers and non-aspirin-takers might be *systematically different*. What does that mean? People who take a daily aspirin are more likely to be more health conscious than non-aspirin-takers. In this instance, the level of health consciousness could be an important Z variable. These individuals likely exercise more and eat a healthier diet. Both of these things, of course, are probably associated with lower blood pressure. What this means is that the comparison between aspirin-takers and non-aspirin-takers is potentially misleading because it is confounded by other factors like diet and exercise. So is the lower blood pressure the result of the aspirin, or is it the result of the better diet and increased exercise that aspirin-takers also benefit from? Because this particular nonexperimental research design does not answer that question, it does not clear our fourth causal hurdle. It is impossible to know whether it was the aspirin that caused the lower blood pressure. In this nonexperimental design just described, because there are other factors that influence blood pressure – and, critically, because these factors are also related to whether or not people take aspirin – it is very difficult to say conclusively that the independent variable (aspirin intake) causes the dependent variable (blood pressure). Figure 4.1 shows this graphically.

Here is where experiments differ so drastically from any other kind of research design. What experimental research designs accomplish by way

[1] Technically, of course, aspirin may or may not cause changes in blood pressure. But even if it does, the evidence just described does not prove it.

of random assignment to treatment groups, then, is to decontaminate the comparison between the treatment and control group of all other influences. Before any stimulus (like a drug or placebo) is administered, all of the participants are in the same pool. Researchers divide them by using some random factor like a coin flip, and that difference is the only difference between the two groups.

Think of it another way. The way that the confounding variables in Figure 4.1 are correlated with the independent variable is highly improbable in an experiment. Why? Because if X is determined by randomness, like a coin flip, then (by the very definition of randomness) it is exceedingly unlikely to be correlated with anything (including confounding variables Z). When researchers control and assign values of X randomly, the comparison between the different groups will not be affected by the fact that other factors certainly do cause Y, the dependent variable.

Connect this back to our discussion from Chapter 3 about how researchers attempt to establish whether some X causes Y. As we will see, experiments are not the only method that help researchers cross the four causal hurdles, but they are uniquely capable in accomplishing that task. Consider each hurdle in turn. First, we should evaluate whether there is a credible causal mechanism before we decide to run the experiment. It is worth noting that the crossing of this causal hurdle is neither easier nor harder in experiments than in nonexperiments. Coming up with a credible causal scenario that links X to Y heightens our dependence on theory, not on data or research design.

Second, in an experiment, it is impossible for Y to cause X – the second causal hurdle – for two reasons. First, assigning X occurs in time before Y is measured, which makes it impossible for Y to cause X. In addition, as previously noted, if X is generated by randomness alone, then nothing (including Y) can cause it.

Establishing, third, whether X and Y are correlated is easily done in any research design (as we will see in Chapter 8). What about our fourth causal hurdle? Is there some Z that is related to both X and Y that makes the association between X and Y spurious? Experiments are uniquely well equipped to help us answer this question definitively. An experiment does not, in any way, eliminate the possibility that a variety of other variables (that we call Z) might also affect Y (as well as X). What the experiment does, through the process of randomly assigning subjects to different values of X, is to equate the treatment and control groups on all possible factors. On every possible variable, whether or not it is related to X, or to Y, or to both, or to neither, the treatment and control groups should, in theory, be identical. That makes the comparison between the two values of X

unpolluted by any possible Z variables because we expect the groups to be equivalent on all values of Z.

In our example of the new drug and blood pressure, by experimentally assigning our participants to treatment (drug) and control (placebo) groups, we do not deny that diet and exercise, for example, might affect blood pressure; we merely neutralize those influences. How? Say that some of our participants are triathletes, who likely have low blood pressure, and others are couch potatoes, who likely have high blood pressure. Randomly assigning all participants – triathletes and couch potatoes alike – to the drug and placebo groups will neutralize the effects of their exercise (or lack thereof) on the aggregate blood-pressure statistics for the two groups. Why? Because, by randomly assigning all participants to the drug or the placebo group, we would expect, on average, half of the triathletes to be in the drug group and half in the placebo group. Likewise, we would expect half of our couch potatoes to be in the treatment group, and half in the placebo group. Thus, when the treatment and control groups' rates of high blood pressure are compared, the effects of different amounts of exercise will not mislead us into thinking that the drug does (or does not) have an effect.

Remarkably, the experimental ability to control for the effects of outside variables (Z) applies to *all* possible confounding variables, regardless of whether we, the researchers, are aware of them. Let's make the example downright preposterous. Let's say that, 20 years from now, another team of scientists discovers that having attached (as opposed to detached) earlobes causes people to have low blood pressure. Does that possibility threaten the inference that we draw from our experiment about our drug and blood pressure today? No, not at all. Why not? Because, whether or not we are aware of it, the random assignment of participants to treatment groups means that, whether we are paying attention to it or not, we would expect our treatment and control groups to have equal numbers of people with attached earlobes, and for both groups to have equal numbers of people with detached earlobes. The key element of an experimental research design – randomly assigning subjects to different values of X, the independent variable – controls for every Z in the universe, whether or not we are aware of that Z.

Together, all of this means that experiments bring with them a particularly strong confidence in the causal inferences drawn from the analysis. In scientific parlance, this is called **internal validity**. If a research design produces high levels of confidence in the conclusions about causality, it is said to have high internal validity. Conversely, research designs that do not allow for particularly definitive conclusions about whether X causes Y are said to have low degrees of internal validity.

4.2.1 "Random Assignment" versus "Random Sampling"

It is critical that you do not confuse the experimental process of randomly assigning people to treatment groups, on the one hand, with the process of randomly sampling people for participation, on the other hand. They are entirely different, and in fact have nothing more in common than that six-letter word "random." They are, however, quite often confused for one another. Random assignment to treatment groups occurs when the participants for an experiment are assigned randomly to one of several possible values of X, the independent variable; importantly, this definition says nothing at all about how the subjects were selected for participation. But random sampling is, at its very heart, about how researchers select people for participation in a study – they are selected at random, that is, every member of the underlying population has an equal probability of being selected. (This is common in survey research, for example.) Mixing up these two critical concepts will produce a good bit of confusion.

4.2.2 Are There Drawbacks to Experimental Research Designs?

Experiments, as we have seen, have a unique ability to get social scientists across our hurdles needed to establish whether X causes Y. But that does not mean they are without their disadvantages. Many of these disadvantages are related to the differences between medical and physical sciences, on the one hand, and the social sciences, on the other. We now discuss four such drawbacks to experimentation.

First, especially in the social sciences, not every independent variable (X) is controllable and subject to experimental manipulation. Suppose, for example, that we wish to study the effects of gender on political participation. Do men contribute more money, vote more, volunteer more in campaigns, than women? There are a variety of nonexperimental ways to study this relationship, but it is impossible to experimentally manipulate a subject's gender. Recall that the definition of an experiment is that the researcher both controls and randomly assigns the values of the independent variable. In this case, the presumed cause (the independent variable) is a person's gender. Compared with drugs versus placebos, assigning a participant's gender is another matter entirely. It is, to put it mildly, impossible. People show up at an experiment either male or female, and it is not within the experimenter's power to "randomly assign" a participant to be male or female.

This is true in many, many political science examples. There are simply myriads of substantive problems that are impossible to study in an experimental fashion. How does a person's partisanship (X) affect his issue

opinions (Y)? How does a person's income level (X) affect her campaign contributions (Y)? How does a country's level of democratization (X) affect its openness to international trade (Y)? How does the level of military spending in India (X) affect the level of military spending in Pakistan (Y) – and, for that matter, vice versa? How does media coverage (X) in an election campaign influence voters' priorities (Y)? In each of these examples that intrigues social scientists, the independent variable is simply not subject to experimental manipulation. Social scientists cannot, in any meaningful sense, "assign" people a party identification or an income, "assign" a country a level of democratization or level of military spending, or "assign" a campaign-specific, long-term amount of media coverage. These variables simply exist in nature, and we cannot control exposure to them and randomly assign different values to different cases (that is, individual people or countries).

A second potential disadvantage of experimental research designs is that experiments often suffer from low degrees of external validity. We have noted that the key strength of experiments is that they typically have high levels of internal validity. That is, we can be quite confident that the conclusions about causality reached in the analysis are not confounded by other variables. External validity, in a sense, is the other side of the coin, as it represents the degree to which we can be confident that the results of our analysis apply not only to the participants in the study, but also to the population more broadly construed. Recall that there is nothing whatsoever in our definition of an experiment that describes how researchers recruit or select people to participate in the experiment. To reiterate: *It is absolutely not the case that experiments require a random sample of the target population.* Indeed, it is extremely rare for experiments to draw a random sample from a population.[2] In drug-trial experiments, for example, it is common to place advertisements in newspapers or on the radio to invite participation, usually involving some form of compensation to the participants. Clearly, people who see and respond to advertisements like this are not a random sample of the population of interest, which is typically thought of as all potential recipients of the drug. Similarly, when professors "recruit" people from their (or their colleagues') classes, the

[2] Since 1990 or so, however, there has been a growing movement in the field of survey research – which has always used random samples of the population – to use computers in the interviewing process that includes experimental randomization of variations in survey questions, in a technique called a "survey experiment." Such designs are intended to reap the benefits of both random assignment to treatment groups, and hence have high internal validity, as well as the benefits of a random sample, and hence have high external validity. See Piazza, Sniderman, and Tetlock (1990) and Sniderman and Piazza (1993).

participants are not a random sample of *any* population.[3] The participant pool in this case represents what we would call a **sample of convenience**, which is to say, "this is more or less the group of people we could beg, coerce, entice, or cajole to participate."

With a sample of convenience, it is simply unclear how, if at all, the results of the experiment generalize to a broader population. As we will learn in Chapter 7, this is a critical issue in the social sciences. Because most experiments make use of such samples of convenience, with any single experiment, it is difficult to know whether the results of that analysis are in any way typical of what we would find in a different sample. With experimental designs, then, scientists learn about how their results apply to a broader population through the process of **replication**, in which researchers implement the same procedures repeatedly in identical form to see if the relationships hold in a consistent fashion.

Experimental research designs, at times, can be plagued with a third disadvantage, namely that they carry special ethical dilemmas for the researcher. Ethical issues about the treatment of human participants occur frequently with medical experiments, of course. If we wished to study experimentally the effects of different types of cancer treatments on survival rates, this would require obtaining a sample of patients with cancer and then randomly assigning the patients to differing treatment regimens. This is typically not considered acceptable medical practice. In such high-stakes medical situations, most individuals value making these decisions themselves, in consultation with their doctor, and would not relinquish the important decisions about their treatment to a random-number generator.

Ethical situations arise less frequently, and typically less dramatically, in social science experimentation, but they do arise on occasion. During the behavioral revolution in psychology in the 1960s, several famous experiments conducted at universities produced vigorous ethical debates. Psychologist Stanley Milgram conducted experiments on how easily he could make individuals obey an authority figure. In this case, the dependent variable was the willingness of the participant to administer what he believed to be a shock to another participant, who was in fact an employee of Milgram's. (The ruse was that Milgram told the participant that he was testing how negative reinforcement – electric shocks – affected the "learning" of the "student.") The independent variable was the degree to which

[3] Think about that for a moment. Experiments in undergraduate psychology or political science classes are not a random sample of 18- to-22 year olds, or even a random sample of undergraduate students, or even a random sample of students from your college or university. Your psychology class is populated with people more interested in the social sciences than in the physical sciences or engineering or the humanities.

Milgram conveyed his status as an authority figure. In other words, the X that Milgram manipulated was the degree to which he presented himself as an authority who must be obeyed. For some participants, Milgram wore a white lab coat and informed them that he was a professor at Yale University. For others, he dressed more casually and never mentioned his institutional affiliation. The dependent variable, then, was how strong the (fake) shocks would be before the subject simply refused to go on. At the highest extreme, the instrument that delivered the "shock" said "450 volts, XXX." The results of the experiment were fascinating because, to his surprise, Milgram found that the great majority of his participants were willing to administer even these extreme shocks to the "learners." But scientific review boards consider such experiments unethical today, because the experiment created a great degree of emotional distress among the true participants.

A fourth potential drawback of experimental research designs is that, when interpreting the results of an experiment, we sometimes make mistakes of emphasis. If an experiment produces a finding that some X does indeed cause Y, that does not mean that that particular X is the most prominent cause of Y. As we have emphasized repeatedly, a variety of independent variables are causally related to every interesting dependent variable in the social sciences. Experimental research designs often do not help to sort out which causes of the dependent variable have the largest effects and which ones have smaller effects.

4.3 OBSERVATIONAL STUDIES (IN TWO FLAVORS)

Taken together, the drawbacks of experiments mean that, for any given political science research situation, implementing an experiment often proves to be unworkable, and sometimes downright impossible. As a result, experimentation is not the most common research design used by political science researchers. In some subfields, such as political psychology – which, as the name implies, studies the cognitive and emotional underpinnings of political decision making – experimentation is quite common. And it is becoming more common in the study of public opinion and electoral competition. But the experiment, for many researchers and for varying reasons, remains a tool that is not applicable to many of the phenomena that we seek to study.

Does this mean that researchers have to shrug their shoulders and abandon their search for causal connections before they even begin? Not at all. But what options do scholars have when they cannot control exposure to different values of the independent variables? In such cases, the only choice is to take the world as it already exists and make the comparison between either individual units – like people, political parties, or countries – or between an aggregated quantity that varies over time. These represent two variants of what is most commonly called an **observational study**. Observational studies are not experiments, but they seek to emulate them. They are known as observational studies because, unlike the controlled and somewhat artificial nature of most experiments, in these research designs, researchers simply take reality as it is and "observe" it, attempting to sort out causal connections without the benefit of randomly assigning participants to treatment groups. Instead, different values of the independent variable already exist in the world, and what scientists do is observe them and then evaluate their theoretical claims by putting them through the same four causal hurdles to discover whether X causes Y.

This leads to the definition of an observational study: An observational study is a research design in which the researcher does *not* have control over values of the independent variable, which occur naturally. However, it is necessary that there be some degree of variability on the independent variable between cases, as well as variation in the dependent variable.

Because there is no random assignment to treatment groups, as in experiments, some scholars claim that it is impossible to speak of causality in observational studies, and therefore sometimes refer to them as **correlational studies**. Along with most political scientists, we do not share this view. Certainly experiments produce higher degrees of confidence about causal matters than do observational studies. However, in observational studies, if sufficient attention is paid to accounting for all of the other possible causes of the dependent variable that are suggested by current understanding, then we can make informed evaluations of their confidence that the independent variable does cause the dependent variable.

Observational studies, as this discussion implies, face exactly the same four causal hurdles as do experiments. (Recall that those hurdles are present in any research design.) So how, in observational studies, do we cross these hurdles? The first causal hurdle – focusing on a credible mechanism connecting X and Y – is identical in experimental and observational studies.

In an observational study, however, crossing the second causal hurdle – can we rule out the possibility that Y causes X? – can sometimes be problematic. For example, do countries with higher levels of economic development (X) have, as a consequence, more stable democratic regimes (Y)? Crossing the second causal hurdle, in this case, is a rather dicey matter. It is clearly plausible that having a stable democratic government makes economic prosperity more likely, which is the reverse-causal scenario. After all, investors are probably more comfortable taking risks with their money in democratic regimes than in autocratic ones. Those risks, in turn, likely produce greater degrees of economic prosperity. It is possible, of course, that X and Y are mutually reinforcing – that is, X causes Y and Y causes X.

The third hurdle – do X and Y covary – is no more difficult for an observational study than for an experiment. (The techniques for examining relationships between two variables are straightforward, and you will learn them in Chapter 8.) But, unlike in an experimental setting, if we fail to find covariation between X and Y in an observational setting, we should still proceed to the fourth hurdle because the possibility remains that we will find covariation between and X and Y when we control for some variable Z (think back to the breast cancer example).

The most pointed comparison between experiments and observational studies, though, occurs with respect to the fourth causal hurdle. The near-magic that happens in experiments because of random assignment to treatment groups – which enables researchers to know that no other factors interfere in the relationship between X and Y – is not present in an observational study. So, in an observational study, the comparison between groups with different values of the independent variable may very well be polluted by other factors, interfering with our ability to make conclusive statements about whether X causes Y.

Within observational studies, there are two pure types – **cross-sectional observational studies**, which focus on variation between **spatial units** for a single **time unit**, and **time-series observational studies**, which focus on variation within a single spatial unit over multiple time units. There are, in addition, hybrid designs, but for sake of simplicity we will focus on the pure types.[4] Before we get into the two types of observational studies, we need to provide a brief introduction to observational data.

4.3.1 Datum, Data, Data Set

The word **"data"** is one of the most grammatically misused words in the English language. Why? Because most people use this word as though it were a singular word when it is, in fact, plural. Any time you read "the data is," you have found a grammatical error. Instead, when describing data, the phrasing should be "the data are." Get used to it: You are now one of the foot soldiers in the crusade to get people to use this word appropriately. It will be a long and uphill battle.

The singular form of the word data is **"datum."** Together, a collection of datum produces data or a "data set." We define observational data sets by the variables that they contain and the spatial and time units over which they are measured. Political scientists use data measured on a variety of

[4] The classic statements of observational studies appeared in 1963 in Donald Campbell and Julian Stanley's seminal work *Experimental and Quasi-experimental Designs for Research*.

Table 4.1. Example of cross-sectional data

Nation	Government debt as a percentage of GNP	Unemployment rate
Finland	6.6	2.6
Denmark	5.7	1.6
United States	27.5	5.6
Spain	13.9	3.2
Sweden	15.9	2.7
Belgium	45.0	2.4
Japan	11.2	1.4
New Zealand	44.6	0.5
Ireland	63.8	5.9
Italy	42.5	4.7
Portugal	6.6	2.1
Norway	28.1	1.7
Netherlands	23.6	2.1
Germany	6.7	0.9
Canada	26.9	6.3
Greece	18.4	2.1
France	8.7	2.8
Switzerland	8.2	0.0
United Kingdom	53.6	3.1
Australia	23.8	2.6

different spatial units. For instance, in survey research, the spatial unit is the individual survey respondent. In comparative U.S. state government studies, the spatial unit is the U.S. state. In international relations, the spatial unit is often the nation. Commonly studied time units are months, quarters, and years. It is also common to refer to the spatial and time units that define data sets as the data set dimensions.

Two of the most common types of data sets correspond directly to the two types of observational studies that we just introduced. For cross-section quasi-experiments, researchers analyze cross-sectional data to determine whether the third causal hurdle has been cleared. For instance, Table 4.1 presents a cross-sectional data set in which the time unit is the year 1972 and the spatial unit is nations. These data could be used to test the theory that unemployment percentage $(X) \rightarrow$ government debt as a percentage of gross national product (Y).

For time-series observational studies, time-series data are analyzed to determine whether the third causal hurdle has been cleared. These data contain measures of X and Y across time for a single spatial unit. For instance, Table 4.2 displays a time-series data set in which the spatial unit is the United States and the time unit is months. We could use these data to

Table 4.2. Example of time-series data

Month	Presidential approval	Inflation
2002.01	83.7	1.14
2002.02	82.0	1.14
2002.03	79.8	1.48
2002.04	76.2	1.64
2002.05	76.3	1.18
2002.06	73.4	1.07
2002.07	71.6	1.46
2002.08	66.5	1.80
2002.09	67.2	1.51
2002.10	65.3	2.03
2002.11	65.5	2.20
2002.12	62.8	2.38

test the theory that inflation $(X) \rightarrow$ presidential approval (Y). In a data set, researchers analyze only those data or data points that contain measured values for both the independent variable (X) and the dependent variable (Y) to determine whether the third causal hurdle has been cleared.

4.3.2 Cross-Sectional Observational Studies

As the name implies, a cross-sectional observational study examines a cross section of social reality, focusing on variation between ***individual spatial units*** – again, like citizens, elected officials, voting districts, or countries – and explaining the variation in the dependent variable across them.

For example, what, if anything, is the connection between the preferences of the voters from a district (X) and a representative's voting behavior (Y)? In a cross-sectional observational study, the strategy that a researcher would pursue in answering this question involves comparing the aggregated preferences of voters from a variety of districts (X) with the voting records of the representatives (Y). Such an analysis, of course, would have to be observational, instead of experimental, because this particular X is not at all subject to experimental manipulation. Such an analysis might take place within the confines of a single legislative session, for a variety of practical purposes (such as the absence of turnover in seats, which is an obviously complicating factor).

Bear in mind, of course, that observational studies have to cross the same four casual hurdles as do experiments. And we have noted that, unlike experiments, with their random assignment to treatment groups, observational studies will often get stuck on our fourth hurdle. That might indeed be the case here. Assuming the other three hurdles can be cleared, consider the possibility that there are confounding variables that cause Y and are also correlated with X, which make the X–Y connection spurious. (Can you think of any such factors?) How do cross-sectional observational studies deal with this critical issue? The answer is that, in most cases, this can be accomplished through a series of rather straightforward statistical controls. In particular, in Chapter 10, you will learn the most common social science research tool for "controlling for" other possible causes of Y, namely the multivariate regression model. What you will learn there is that multivariate regression can allow researchers to see how, if at all, controlling for another variable (like *Z*) affects the relationship between *X* and *Y*.

4.3.3 Time-Series Observational Studies

The other major variant of observational studies is the time-series observational study, which has, at its heart, a comparison over time within a single spatial unit. Unlike in the cross-sectional variety, which examines relationships between variables across individual units typically at a single time point, in the time-series observational study, political scientists typically examine the variation within one spatial unit over time.[5]

For example, how, if at all, do changes in media coverage about the economy (*X*) affect public concern about the economy (*Y*)?[6] To be a bit more specific, when the media spend more time talking about the potential problem of inflation, does the public show more concern about inflation, and when the media spend less time on the subject of inflation, does public concern about inflation wane? We can measure these variables in aggregate terms that vary over time. For example, how many stories about inflation make it onto the nightly news in a given month? It is almost certain ~~that~~ that quantity will not be the same each and every month. And how much concern does the public show (through opinion polls, for example) about inflation in a given month? Again, the percentage of people who identify inflation as a pressing problem will almost certainly vary from month to month.

Of course, as with its cross-sectional cousin, the time-series observational study will require us to focus hard on that fourth causal hurdle. Are there any other variables (*Z*) that are related to the varying volume of news coverage about inflation (*X*) and public concern about inflation (*Y*)? (The third exercise at the end of this chapter will ask for your thoughts on this subject.) If we can identify any other possible causes of why the public is

[5] The spatial units analyzed in time-series observational studies are usually aggregated.
[6] See Iyengar and Kinder (1987).

sometimes more concerned about inflation, and why they are sometimes less concerned about it, then we will need to control for those factors in our analysis.

4.3.4 The Major Difficulty with Observational Studies

We noted that experimental research designs carry some drawbacks with them. So, too, do observational studies. Here, we focus only on one, but it is a big one. As the preceding examples demonstrate, when we need to control for the other possible causes of *Y* to cross the fourth causal hurdle, we need to control for *all of them*, not just one.[7] But how do we know whether we have controlled for all of the other possible causes of *Y*? In many cases, we don't know that for certain. We need to try, of course, to control statistically for all other possible causes that we can, which involves carefully considering the previous research on the subject and gathering as much data on those other causes as is possible. But in many cases, we will simply be unable to do this perfectly.

What all of this means, in our view, is that observational analysis must be a bit more tentative in its pronouncements about causality. Indeed, if we have done the very best we can to control for as many causes of *Y*, then the most sensible conclusion we can reach, in many cases, is that *X* causes *Y*. But in practice, our conclusions are rarely definitive, and subsequent research can modify them. That can be frustrating, we know, for students to come to grips with – and it can be frustrating for researchers, too. But the fact that conclusive answers are difficult to come by should only make us work harder to identify other causes of *Y*.

4.4 SUMMARY

For almost every phenomenon of interest to political scientists, there is more than one form of research design that they could implement to address questions of causal relationships. Before starting a project, researchers need to decide whether to use experimental or observational methods; and if they opt for the latter, as is common, they have to decide what type of observational study to use. And sometimes researchers choose more than one type of design.

Different research designs help shed light on different questions. Focus, for the moment, on a simple matter like preferences for a more liberal

[7] As we will see in Chapter 10, technically we need to control only for the factors that might affect *Y* and are also related to *X*. In practice, though, that is a very difficult distinction to make.

or conservative government policy. Cross-sectional and time-series approaches are both useful in this respect. They simply address different types of substantive questions. Cross-sectional approaches look to see why some other individuals prefer more liberal government policies, and why some individuals prefer more conservative government policies. That is a perfectly worthwhile undertaking for a political scientist: What causes some people to be liberals and others to be conservatives? But consider the time-series approach, which focuses on why the public as an aggregated whole prefers a more liberal or a more conservative government at different points in time. That is simply a different question. Neither approach is inherently better or worse than the other, but they both shed light on different aspects of social reality. Which design researchers should choose depends on what type of question they intend to ask and answer.

CONCEPTS INTRODUCED IN THIS CHAPTER

- aggregate
- control group
- correlational studies
- cross-sectional observational studies
- data
- data points
- data set dimensions
- datum
- experiment
- external validity
- internal validity
- observational study
- random assignment to treatment groups
- random sampling
- replication
- research designs
- sample of convenience
- spatial units
- time units
- time-series observational studies
- treatment group

EXERCISES

1. Consider the following proposed relationships between an independent and a dependent variable. In each case, would it be realistic for a researcher to perform an experiment to test the theory? If yes, briefly describe what would be randomly assigned in the experiment; if not, briefly explain why not.
   (a) An individual's level of religiosity (X) and his or her preferences for different political candidates (Y)
   (b) Exposure to negative political news (X) and political apathy (Y)
   (c) Military service (X) and attitudes toward foreign policy (Y)
   (d) A speaker's personal characteristics (X) and persuasiveness (Y)
2. Consider the relationship between education level (X) and voting turnout (Y). How would the design of a cross-sectional observational study differ from that of a time-series observational study?

3. In the section on time-series observational studies, we introduced the idea of how varying levels of media coverage of inflation (X) might cause variation in public concern about inflation (Y). Can you think of any relevant Z variables that we will need to control for, statistically, in such an analysis, to be confident that the relationship between X and Y is causal?
4. In the previous chapter (specifically, Sections 3.2 and 3.3), we gave examples of research problems. For each of these examples, identify the spatial unit(s) and time unit(s). For each, say whether the study was an experiment, a cross-sectional observational study, or a time-series observational study.
5. Table 4.1 presents data for a test of a theory by use of a cross-sectional observational study. If this same theory were tested by use of a time-series observational study, what would the data table look like?
6. Compare the two designs for testing the preceding theory. Across the two forms of observational studies, what are the Z variables for which you want to control?
7. Table 4.2 presents data for a test of a theory by use of a time-series observational study. If this same theory were tested by use of a cross-sectional observational study, what would the data table look like?
8. Compare the two designs for testing the preceding theory. Across the two forms of observational studies, what are the Z variables for which you want to control?