
Designing Social Inquiry

SCIENTIFIC INFERENCE IN
QUALITATIVE RESEARCH

Gary King

Robert O. Keohane

Sidney Verba

PRINCETON UNIVERSITY PRESS

PRINCETON, NEW JERSEY

Chapter 3, Section 3.5: Rules for Constructing Causal Theories

3.5 RULES FOR CONSTRUCTING CAUSAL THEORIES

Much sensible advice about improving qualitative research is precise, specific, and detailed; it involves a manageable and therefore narrow aspect of qualitative research. However, even in the midst of solving a host of individual problems, we must keep the big picture firmly in mind: each specific solution must help in solving whatever is the general causal inference problem one aims to solve. Thus far in this chapter, we have provided a precise theoretical definition of a causal effect and discussed some of the issues involved in making causal inferences. We take a step back now and provide a broader overview of some rules regarding theory construction. As we discuss (and have discussed in section 1.2), improving theory does not end when data collection begins.

Causal theories are designed to show the causes of a phenomenon or set of phenomena. Whether originally conceived as deductive or inductive, any theory includes an interrelated set of causal hypotheses. Each hypothesis specifies a posited relationship between variables that creates observable implications: if the specified explanatory variables

take on certain values, other specified values are predicted for the dependent variables. Testing or evaluating any causal hypothesis requires causal inference. The overall theory, of which the hypotheses are parts should be *internally consistent*, or else hypotheses can be generated that contradict one another.

Theories and hypotheses that fit these definitions have an enormous range. In this section, we provide five rules that will help in formulating good theories, and we provide a discussion of each with examples.

3.5.1 Rule 1: Construct Falsifiable Theories

By this first rule, we do not only mean that a “theory” incapable of being wrong is not a theory. We also mean that we should design theories so that they can be shown to be wrong as easily and quickly as possible. Obviously, we should not actually try to be wrong, but even an incorrect theory is better than a statement that is neither wrong nor right. The emphasis on falsifiable theories forces us to keep the right perspective on uncertainty and guarantees that we treat theories as tentative and not let them become dogma. We should always be prepared to reject theories in the face of sufficient scientific evidence against them. One question that should be asked about any theory (or of any hypothesis derived from the theory) is simply: what evidence would falsify it? The question should be asked of all theories and hypotheses but, above all, the researcher who poses the theory in the first place should ask it of his or her own.

Karl Popper is most closely identified with the idea of falsifiability (Popper 1968). In Popper’s view, a fundamental asymmetry exists between confirming a theory (verification) and disconfirming it (falsification). The former is almost irrelevant, whereas the latter is the key to science. Popper believes that a theory once stated immediately becomes part of the body of accepted scientific knowledge. Since theories are general, and hypotheses specific, theories technically imply an infinite number of hypotheses. However, empirical tests can only be conducted on a finite number of hypotheses. In that sense, “theories are not verifiable” because we can never test all observable implications of a theory (Popper 1968:252). Each hypothesis tested may be shown to be consistent with the theory, but any number of consistent empirical results will not change our opinions since the theory remains accepted scientific knowledge. On the other hand, if even a single hypothesis is shown to be wrong, and thus inconsistent with the theory, the theory is falsified, and it is removed from our collection of scientific knowledge. “The passing of tests therefore makes not a jot of difference to the status of any hypothesis, though the failing of just one test may

make a great deal of difference" (Miller 1988:22). Popper did not mean falsification to be a deterministic concept. He recognized that any empirical inference is to some extent uncertain (Popper 1982). In his discussion of disconfirmation, he wrote, "even if the asymmetry [between falsification and verification] is admitted, it is still impossible, for various reasons, that any theoretical system should ever be conclusively falsified" (Popper 1968:42).

In our view, Popper's ideas are fundamental for *formulating* theories. We should always design theories that are vulnerable to falsification. We should also learn from Popper's emphasis on the tentative nature of any theory. However, for *evaluating* existing social scientific theories, the asymmetry between verification and falsification is not as significant. Either one adds to our scientific knowledge. The question is less whether, in some general sense, a theory is false or not—virtually every interesting social science theory has at least one observable implication that appears wrong—than *how much of the world the theory can help us explain*. By Popper's rule, theories based on the assumption of rational choice would have been rejected long ago since they have been falsified in many specific instances. However, social scientists often choose to retain the assumption, suitably modified, because it provides considerable power in many kinds of research problems (see Cook and Levi 1990). The same point applies to virtually every other social science theory of interest. The process of trying to falsify theories in the social sciences is really one of searching for their bounds of applicability. If some observable implication indicates that the theory does not apply, we learn something; similarly, if the theory works, we learn something too.

For scientists (and especially for social scientists) evaluating properly formulated theories, Popper's fundamental asymmetry seems largely irrelevant. O'Hear (1989:43) made a similar point about the application of Popper's ideas to the physical sciences:

Popper always tends to speak in terms of *explanations* of *universal* theories. But once again, we have to insist that proposing and testing universal theories is only part of the aim of science. There may be no true universal theories, owing to conditions differing markedly through time and space; this is a possibility we cannot overlook. But even if this were so, science could still fulfil [sic] many of its aims in giving us knowledge and true predictions about conditions in and around our spatio-temporal niche.

Surely this same point applies even more strongly to the social sciences.

Furthermore, Popper's evaluation of theories does not fundamentally distinguish between a newly formulated theory and one that has

withstood numerous empirical tests. When we are testing for the deterministic distinction between the truth or fiction of a universal theory (of which there exists no interesting examples), Popper's view is appropriate, but from our perspective of searching for the bounds of a theory's applicability, his view is less useful. As we have indicated many times in this book, we require all inferences about specific hypotheses to be made by stating a best guess (an estimate) and a measure of the uncertainty of this guess. Whether we discover that the inference is consistent with our theory or inconsistent, our conclusion will have as much effect on our belief in the theory. Both consistency and inconsistency provide information about the truth of the theory and should affect the certainty of our beliefs.¹³

Consider the hypothesis that Democratic and Republican campaign strategies during American presidential elections have a small net effect on the election outcome. Numerous more specific hypotheses are implied by this one, such as that television commercials, radio commercials, and debates all have little effect on voters. Any test of the theory must really be a test of one of these hypotheses. One test of the theory has shown that forecasts of the outcome can be made very accurately with variables available only at the time of the conventions—and thus before the campaign (Gelman and King 1993). This test is consistent with the theory (if we can predict the election before the campaign, the campaign can hardly be said to have much of an impact), but it does not absolutely verify it. Some aspect of the campaign could have some small effect that accounts for some of the forecasting errors (and few researchers doubt that this is true). Moreover, the prediction could have been luck, or the campaign could have not included any innovative (and hence unpredictable) tactics during the years for which data were collected.

We could conduct numerous other tests by including variables in the forecasting model that measure aspects of the campaign, such as relative amounts of TV and radio time, speaking ability of the candidates, and judgements as to the outcomes of the debates. If all of these hypotheses show no effect, then Popper would say that our opinion is not changed in any interesting way: the theory that presidential campaigns have no effect is still standing. Indeed, if we did a thousand

¹³ Some might call us (or accuse us of being!) "justificationists" or even "probabilistic justificationists" (see Lakatos 1970), but if we must be labeled, we prefer the more coherent, philosophical Bayesian label (see Leamer 1978; Zellner 1971; and Barnett 1982). In fact, our main difference with Popper is our goals. Given his precise goal, we agree with his procedure; given our goal, perhaps he might agree with ours. However, we believe that our goals are closer to those in use in the social sciences and are also closer to the ones likely to be successful.

similar tests and all were consistent with the theory, the theory could still be wrong since we have not tried every one of the infinite number of possible variables measuring the campaign. So even with a lot of results consistent with the theory, it still *might* be true that presidential campaigns influence voter behavior.

However, if a single campaign event—such as substantial accusations of immoral behavior—is shown to have some effect on voters, the theory would be falsified. According to Popper, even though this theory was not conclusively falsified (which he recognized as impossible), we learn more from it than the thousand tests consistent with the theory.

To us, this is not the way social science is or should be conducted. After a thousand tests in favor and one against, even if the negative test seemed valid with a high degree of certainty, we would not drop the theory that campaigns have no effect. Instead, we might modify it to say perhaps that normal campaigns have no effect except when there is considerable evidence of immoral behavior by one of the candidates—but since this modification would make our theory more restrictive, we would need to evaluate it with a new set of data before being confident of its validity. The theory would still be very powerful, and we would know somewhat more about the bounds to which the theory applied with each passing empirical evaluation. Each test of a theory affects both the estimate of its validity and the uncertainty of that estimate; and it may also affect to what extent we wish the theory to apply.

In the previous discussion, we suggested an important approach to theory, as well as issued a caution. The approach we recommended is one of sensitivity to the contingent nature of theories and hypotheses. Below, we argue for seeking broad application for our theories and hypotheses. This is a useful research strategy, but we ought always to remember that theories in the social sciences are unlikely to be universal in their applicability. Those theories that are put forward as applying to everything, everywhere—some versions of Marxism and rational choice theory are examples of theories that have been put forward with claims of such universality—are either presented in a tautological manner (in which case they are neither true nor false) or in a way that allows empirical disconfirmation (in which case we will find that they make incorrect predictions). Most useful social science theories are valid under particular conditions (in election campaigns without strong evidence of immoral behavior by a candidate) or in particular settings (in industrialized but not less industrialized nations, in House but not Senate campaigns). We should always try to specify the bounds of applicability of the theory or hypothesis. The next step is to

raise the question: Why do these bounds exist? What is it about Senate races that invalidates generalizations that are true for House races? What is it about industrialization that changes the causal effects? What variable is missing from our analysis which could produce a more generally applicable theory? By asking such questions, we move beyond the boundaries of our theory or hypothesis to show what factors need to be considered to expand its scope.

But a note of caution must be added. We have suggested that the process of evaluating theories and hypotheses is a flexible one: particular empirical tests neither confirm nor disconfirm them once and for all. When an empirical test is inconsistent with our theoretically based expectations, we do not immediately throw out the theory. We may do various things: We may conclude that the evidence may have been poor due to chance alone; we may adjust what we consider to be the range of applicability of a theory or hypothesis even if it does not hold in a particular case and, through that adjustment, maintain our acceptance of the theory or hypothesis. Science proceeds by such adjustments; but they can be dangerous. If we take them too far we make our theories and hypotheses invulnerable to disconfirmation. The lesson is that we must be very careful in adapting theories to be consistent with new evidence. We must avoid stretching the theory beyond all plausibility by adding numerous exceptions and special cases.

If our study disconfirms some aspect of a theory, we may choose to retain the theory but add an exception. Such a procedure is acceptable as long as we recognize the fact that we are reducing the claims we make for the theory. The theory, though, is less valuable since it explains less; in our terminology, we have less *leverage* over the problem we seek to understand.¹⁴ Furthermore, such an approach may yield a “theory” that is merely a useless hodgepodge of various exceptions and exclusions. At some point we must be willing to discard theories and hypotheses entirely. Too many exceptions, and the theory should be rejected. Thus, by itself, *parsimony, the normative preference for theories with fewer parts, is not generally applicable*. All we need is our more general notion of maximizing leverage, from which the idea of parsimony can be fully derived when it is useful. The idea that science is largely a process of explaining many phenomena with just a few makes clear that theories with fewer parts are not better or worse. To maximize leverage, we should attempt to formulate theories that explain as much as possible with as little as possible. Sometimes this formulation is achieved via parsimony, but sometimes not. We can con-

¹⁴ As always, when we do modify a theory to be consistent with evidence we have collected, then the theory (or that part of it on which our evidence bears) should be evaluated in a different context or new data set.

ceive of examples by which a slightly more complicated theory will explain vastly more of the world. In such a situation, we would surely use the nonparsimonious theory, since it maximizes leverage more than the more parsimonious theory.¹⁵

3.5.2 Rule 2: Build Theories That Are Internally Consistent

A theory which is internally inconsistent is not only falsifiable—it is false. Indeed, this is the only situation where the veracity of a theory is known without any empirical evidence: if two or more parts of a theory generate hypotheses that contradict one another, then no evidence from the empirical world can uphold the theory. Ensuring that theories are internally consistent should be entirely uncontroversial, but consistency is frequently difficult to achieve. One method of producing internally consistent theories is with formal, mathematical modeling. *Formal modeling* is a practice most developed in economics but increasingly common in sociology, psychology, political science, anthropology, and elsewhere (see Ordeshook 1986). In political science, scholars have built numerous substantive theories from mathematical models in rational choice, social choice, spatial models of elections, public economics, and game theory. This research has produced many important results, and large numbers of plausible hypotheses. One of the most important contributions of formal modeling is revealing the internal inconsistency in verbally stated theories.

However, as with other hypotheses, formal models do not constitute verified explanations without empirical evaluation of their predic-

¹⁵ Another formulation of Popper's view is that "you can't prove a negative." You cannot, he argues, because a result consistent with the hypothesis might just mean that you did the wrong test. Those who try to prove the negative will always run into this problem. Indeed, their troubles will be not only theoretical but professional as well since journals are more likely to publish positive results rather than negative ones.

This has led to what is called *the file drawer problem*, which is clearest in the quantitative literature. Suppose no patterns exist in the world. Then five of every one hundred tests of any pattern will fall outside the 95 percent confidence interval and thus produce incorrect inferences. If we were to assume that journals publish positive rather than negative results, they will publish only those 5 percent that are "significant"; that is, they will publish only the papers that come to the wrong conclusions, and our file drawers will be filled with all the papers that come to the right conclusions! (See Iyengar and Greenhouse (1988) for a review of the statistical literature on this problem.) In fact, these incentives are well known by researchers, and it probably affects their behaviors as well. Even though the acceptance rate at many major social science journals is roughly 5 percent, the situation is not quite this bad, but it is still a serious problem. In our view, the file drawer problem could be solved if everyone adopted our alternative position. *A negative result is as useful as a positive one; both can provide just as much information about the world.* So long as we present our estimates and a measure of our uncertainty, we will be on safe ground.

tions. Formality does help us reason more clearly, and it certainly ensures that our ideas are internally consistent, but it does not resolve issues of empirical evaluation of social science theories. An assumption in a formal model in the social sciences is generally a convenience for mathematical simplicity or for ensuring that an equilibrium can be found. Few believe that the political world is mathematical in the same way that some physicists believe the physical world is. Thus, formal models are merely models—abstractions that should be distinguished from the world we study. Indeed, some formal theories make predictions that depend on assumptions that are vastly oversimplified, and these theories are sometimes not of much empirical value. They are only more precise in the abstract than are informal social science theories: they do not make more specific predictions about the real world, since the conditions they specify do not correspond, even approximately, to actual conditions.

Simplifications are essential in formal modeling, as they are in all research, but we need to be cautious about the inferences we can draw about reality from the models. For example, assuming that all omitted variables have no effect on the results can be very useful in modeling. In many of the formal models of qualitative research that we present throughout this book, we do precisely this. Assumptions like this are not usually justified as a feature of the world; they are only offered as a convenient feature of our model of the world. The results, then, apply exactly to the situation in which these omitted variables are irrelevant and may or may not be similar to results in the real world. We do not have to check the assumption to work out the model and its implications, but it is *essential* that we check the assumption during empirical evaluation. The assumption need not be correct for the formal model to be useful. But we cannot take untested or unjustified theoretical assumptions and use them in constructing empirical research designs. Instead, we must generally supplement a formal theory with additional features to make it useful for empirical study.

A good formal model should be abstract so that the key features of the problem can be apparent and mathematical reasoning can be easily applied. Consider, then, a formal model of the effect of proportional representation on political party systems, which implies that proportional representation fragments party systems. The key causal variable is the type of electoral system—whether it is a proportional representation system with seats allocated to parties on the basis of their proportion of the vote or a single-member district system in which a single winner is elected in each district. The dependent variable is the number of political parties, often referred to as the degree of party-system fragmentation. The leading hypothesis is that electoral systems

based on proportional representation generate more political parties than do district-based electoral systems. For the sake of simplicity, such a model might well include only variables measuring some essential features of the electoral system and the degree of party-system fragmentation. Such a model would generate only a *hypothesis*, not a conclusion, about the relationship between proportional representation and party-system fragmentation in the real world. Such a hypothesis would have to be tested through the use of qualitative or quantitative empirical methods.

However, even though an implication of this model is that proportional representation fragments political parties, and even though no other variables were used in the model, using only two variables in an empirical analysis would be foolish. A study that indicates that countries with proportional representation have more fragmented party systems would ignore the problem of endogeneity (section 5.4), since countries which establish electoral systems based on a proportional allocation of seats to the parties may well have done so because of their already existent fragmented party systems. Omitted variable bias would also be a problem since countries with deep racial, ethnic, or religious divisions are probably also likely to have fragmented party systems, and countries with divisions of these kinds are more likely to have proportional representation.

Thus, both of the requirements for omitted variable bias (section 5.2) seem to be met: the omitted variable is correlated both with the explanatory and the dependent variable, and any analysis ignoring the variable of social division would therefore produce biased inferences.

The point should be clear: formal models are extremely useful for clarifying our thinking and developing internally consistent theories. For many theories, especially complex, verbally stated theories, it may be that only a formal model is capable of revealing and correcting internal inconsistencies. At the same time, formal models are unlikely to provide the correct empirical model for empirical testing. They certainly do not enable us to avoid any of the empirical problems of scientific inference.

3.5.3 Rule 3: Select Dependent Variables Carefully

Of course, we should do everything in research carefully, but choosing variables, especially dependent variables, is a particularly important decision. We offer the following three suggestions (based on mistakes that occur all too frequently in the quantitative and qualitative literatures):

First, *dependent variables should be dependent*. A very common mistake is to choose a dependent variable which in fact causes changes in our

explanatory variables. We analyze the specific consequences of endogeneity and some ways to circumvent the problem in section 5.4, but we emphasize it here because the easiest way to avoid it is to choose explanatory variables that are clearly exogenous and dependent variables that are endogenous.

Second, *do not select observations based on the dependent variable so that the dependent variable is constant*. This, too, may seem a bit obvious, but scholars often choose observations in which the dependent variable does not vary at all (such as in the example discussed in section 4.3.1). Even if we do not deliberately design research so that the dependent variable is constant, it may turn out that way. But, as long as we have not predetermined that fact by our selection criteria, there is no problem. For example, suppose we select observations in two categories of an explanatory variable, and the dependent variable turns out to be constant across the two groups. This is merely a case where the estimated causal effect is zero.

Finally we should *choose a dependent variable that represents the variation we wish to explain*. Although this point seems obvious, it is actually quite subtle, as illustrated by Stanley Lieberman (1985:100):

A simple gravitational exhibit at the Ontario Science Centre in Toronto inspires a heuristic example. In the exhibit, a coin and a feather are both released from the top of a vacuum tube and reach the bottom at virtually the same time. Since the vacuum is not a total one, presumably the coin reaches the bottom slightly ahead of the feather. At any rate, suppose we visualize a study in which a variety of objects is dropped without the benefit of such a strong control as a vacuum—just as would occur in nonexperimental social research. If social researchers find that the objects differ in the time that they take to reach the ground, typically they will want to know what characteristics determine these differences. Probably such characteristics of the objects as their density and shape will affect speed of the fall in a nonvacuum situation. If the social researcher is fortunate, such factors together will fully account for all of the differences among the objects in the velocity of their fall. If so, the social researcher will be very happy because all of the variation between objects will be accounted for. The investigator, applying standard social research-thinking will conclude that there is a complete understanding of the phenomenon *because all differences among the objects under study have been accounted for*. Surely there must be something faulty with our procedures if we can approach such a problem without even considering gravity itself.

The investigator's procedures in this example would be faulty only if the variable of interest were gravity. If gravity were the explanatory variable we cared about, our experiment does not vary it (since the

experiment takes place in only one location) and therefore tells us nothing about it. However, the experiment Lieberman describes would be of great interest if we sought to understand variations in the time it will take for different types of objects to hit the ground when they are dropped from the same height under different conditions of air pressure. Indeed, even if we knew all about gravity, this experiment would still yield valuable information. But if, as Lieberman assumes, we were really interested in an inference about the causal effect of gravity, we would need a dependent variable which varied over observations with differing degrees of gravitational attraction. Likewise, in social science, we must be careful to ensure that we are really interested in understanding our dependent variable, rather than the background factors that our research design holds constant.

Thus, we need the entire range of variation in the dependent variable to be a possible outcome of the experiment in order to obtain an unbiased estimate of the impact of the explanatory variables. Artificial limits on the range or values of the dependent variable produce what we define (in section 4.3) as selection bias. For instance, if we are interested in the conditions under which armed conflict breaks out, we cannot choose as observations only those instances where the result is armed conflict. Such a study might tell us a great deal about variations among observations of armed conflict (as the gravity experiment tells us about variations in speed of fall of various objects) but will not enable us to explore the sources of armed conflict. A better design if we want to understand the sources of armed conflict would be one that selected observations according to our explanatory variables and allowed the dependent variable the *possibility* of covering the full range from there being little or no threat of a conflict through threat situations to actual conflict.

3.5.4 Rule 4: Maximize Concreteness

Our fourth rule, which follows from our emphasis on falsifiability, consistency, and variation in the dependent variable is to maximize concreteness. We should choose observable, rather than unobservable, concepts wherever possible. Abstract, unobserved concepts such as utility, culture, intentions, motivations, identification, intelligence, or the national interest are often used in social science theories. They can play a useful role in theory *formulation*; but they can be a hindrance to empirical *evaluation* of theories and hypotheses unless they can be defined in a way such that they, or at least their implications, can be observed and measured. Explanations involving concepts such as culture or national interest or utility or motivation are suspect unless we can

measure the concept independently of the dependent variable that we are explaining. When such terms are used in explanations, it is too easy to use them in ways that are tautological or have no differentiating, observable implications. An act of an individual or a nation may be explained as resulting from a desire to maximize utility, to fulfill intentions, or to achieve the national interest. But the evidence that the act maximized utility or fulfilled intentions or achieved the national interest is the fact that the actor or the nation engaged in it. It is incumbent upon the researcher formulating the theory to specify clearly and precisely what observable implications of the theory would indicate its veracity and distinguish it from logical alternatives.

In no way do we mean to imply by this rule that concepts like intentions and motivations are unimportant. We only wish to recognize that the standard for explanation in any *empirical* science like ours must be *empirical* verification or falsification. Attempting to find empirical evidence of abstract, unmeasurable, and unobservable concepts will necessarily prove more difficult and less successful than for many imperfectly conceived specific and concrete concepts. The more abstract our concepts, the less clear will be the observable consequences and the less amenable the theory will be to falsification.

Researchers often use the following strategy. They begin with an abstract concept of the sort listed above. They agree that it cannot be measured directly; therefore, they suggest specific indicators of the abstract concept that can be measured and use them in their explanations. The choice of the specific indicator of the more abstract concept is justified on the grounds that it is observable. Sometimes it is the *only thing* that is observable (for instance, it is the only phenomenon for which data are available or the only type of historical event for which records have been kept). This is a perfectly respectable, indeed usually necessary, aspect of empirical investigation.

Sometimes, however, it has an unfortunate side. Often the specific indicator is far from the original concept and has only an indirect and uncertain relationship to it. It may not be a valid indicator of the abstract concept at all. But, after a quick apology for the gap between the abstract concept and the specific indicator, the researcher labels the indicator with the abstract concept and proceeds onward as if he were measuring that concept directly. Unfortunately, such reification is common in social science work, perhaps more frequently in quantitative than in qualitative research, but all too common in both. For example, the researcher has figures on mail, trade, tourism and student exchanges and uses these to compile an index of "societal integration" in Europe. Or the researcher asks some survey questions as to whether

respondents are more concerned with the environment or making money and labels different respondents as “materialists” and “post-materialists.” Or the researcher observes that federal agencies differ in the average length of employment of their workers and converts this into a measure of the “institutionalization” of the agencies.

We should be clear about what we mean here. The gap between concept and indicator is inevitable in much social science work. And we use general terms rather than specific ones for good reasons: they allow us to expand our frame of reference and the applicability of our theories. Thus we may talk of legislatures rather than of more narrowly defined legislative categories such as parliaments or specific institutions such as the German Bundestag. Or we may talk of “decision-making bodies” rather than legislatures when we want our theory to apply to an even wider range of institutions. (In the next section we, in fact, recommend this.) Science depends on such abstract classifications—or else we revert to summarizing historical detail. But our abstract and general terms must be connected to specific measureable concepts at some point to allow empirical testing. The fact of that connection—and the distance that must be traversed to make it—must always be kept in mind and made explicit. Furthermore, the choice of a high level of abstraction must have a real justification in terms of the theoretical problem at hand. It must help make the connection between the specific research at hand—in which the particular indicator is the main actor—and the more general problem. And it puts a burden on us to see that additional research using other specific indicators is carried on to bolster the assumption that our specific indicators really relate to some broader concept. The abstract terms used in the examples above—“societal integration,” “post-materialism,” and “institutionalization”—may be measured reasonably by the specific indicators cited. We do not deny that the leap from specific indicator to general abstract concept must be made—we have to make such a leap to carry on social science research. The leap must, however, be made with care, with justification, and with a constant “memory” of where the leap began.

Thus, we do not argue against abstractions. But we do argue for a language of social research that is as concrete and precise as possible. If we have no alternative to using unobservable constructs, as is usually the case in the social sciences, then we should at least *choose ideas with observable consequences*. For example, “intelligence” has never been directly observed but it is nevertheless a very useful concept. We have numerous tests and other ways to evaluate the implications of intelligence. On the other hand, if we have the choice between “the institu-

tionalization of the presidency” and “size of the White House staff,” it is usually better to choose the latter. We may argue that the size of the White House staff is related to the general concept of the institutionalization of the presidency, but we ought not to reify the narrower concept as identical to the broader. And, if size of staff means institutionalization, we should be able to find other measures of institutionalization that respond to the same explanatory variables as does size of staff. Below, we shall discuss “maximizing leverage” by expanding our dependent variables.

Our call for concreteness extends, in general, to the words we use to describe our theory. If a reader has to spend a lot of time extracting the precise meanings of the theory, the theory is of less use. There should be as little controversy as possible over what we mean when we describe a theory. To help in this goal of specificity, even if we are not conducting empirical research ourselves, we should spend time explicitly considering the observable implications of the theory and even possible research projects we could conduct. The vaguer our language, the less chance we will be wrong—but the less chance our work will be at all useful. It is better to be wrong than vague.

In our view, eloquent writing—a scarce commodity in social science—should be encouraged (and savored) in presenting the rationale for a research project, arguing for its significance, and providing rich descriptions of events. Tedium never advanced any science. However, as soon as the subject becomes causal or descriptive inference, where we are interested in observations and generalizations that are expected to persist, we require concreteness and specificity in language and thought.¹⁶

¹⁶ The rules governing the best questions to ask in interviews are almost the same as those used in designing explanations: Be as concrete as possible. We should not ask conservative, white Americans, “Are you racist?”, rather, “Would you mind if your daughter married a black man?” We should not ask someone if he or she is knowledgeable about politics; we should ask for the names of the Secretary of State and Speaker of the House. In general and wherever possible, *we must not ask an interviewee to do our work for us*. It is best not to ask for estimates of causal effects; we must ask for measures of the explanatory and dependent variables, and estimate the causal effect ourselves. We must not ask for motivations, but rather for facts.

This rule is not meant to imply that we should never ask people why they did something. Indeed, asking about motivations is often a productive means of generating hypotheses. Self-reported motivations may also be a useful set of observable implications. However, the answer given must be interpreted as the interviewee’s response to the researcher’s question, not necessarily as the correct answer. If questions such as these are to be of use, we should design research so that a particular answer given (with whatever justifications, embellishments, lies, or selective memories we may encounter) is an observable implication.

3.5.5 Rule 5: State Theories in as Encompassing Ways as Feasible

Within the constraints of guaranteeing that the theory will be falsifiable and that we maximize concreteness, the theory should be formulated so that it explains as much of the world as possible. We realize that there is some tension between this fifth rule and our earlier injunction to be concrete. We can only say that both goals are important, though in many cases they may conflict, and we need to be sensitive to both in order to draw a balance.

For example, we must not present our theory as if it only applies to the German Bundestag when there is reason to believe that it might apply to all independent legislatures. We need not provide evidence for all implications of the theory in order to state it, so long as we provide a reasonable estimate of uncertainty that goes along with it. It may be that we have provided strong evidence in favor of the theory in the German Bundestag. Although we have no evidence that it works elsewhere, we have no evidence against it either. The broader reference is useful if we remain aware of the need to evaluate its applicability. Indeed, expressing it as a hypothetically broader reference may force us to think about the structural features of the theory that would make it apply or not to other independent legislatures. For example, would it apply to the U.S. Senate, where terms are staggered, to the New Hampshire Assembly, which is much larger relative to the number of constituents, or to the British House of Commons, in which party voting is much stronger? An important exercise is stating what we think are systematic features of the theory that make it applicable in different areas. We may learn that we were wrong, but that is considerably better than not having stated the theory with sufficient precision in the first place.

This rule might seem to conflict with Robert Merton's ([1949] 1968) preference for "theories of the middle-range," but even a cursory reading of Merton should indicate that this is not so. Merton was reacting to a tradition in sociology where "theories" such as Parson's "theory of action" were stated so broadly that they could not be falsified. In political science, Easton's "systems theory" (1965) is in this same tradition (see Eckstein 1975:90). As one example of the sort of criticism he was fond of making, Merton ([1949] 1968: 43) wrote, "So far as one can tell, the theory of role-sets is not inconsistent with such broad theoretical orientations as Marxist theory, functional analysis, social behaviorism, Sorokin's integral sociology, or Parson's theory of action." Merton is not critical of the theory of role-sets, which he called a middle-range theory, rather he is arguing against those "broad theoretical orienta-

tions," with which almost any more specific theory or empirical observation is consistent. Merton favors "middle-range" theories but we believe he would agree that theories should be stated as broadly as possible as long as they remain falsifiable and concrete. Stating theories as broadly as possible is, to return to a notion raised earlier, a way of maximizing leverage. If the theory is testable—and the danger of very broad theories is, of course, that they may be phrased in ways that are not testable—then the broader the better; that is, the broader, the greater the leverage.

Chapter 4, Section 4.3 - Selection Bias

4.3 SELECTION BIAS

How should we select observations for inclusion in a study? If we are interviewing city officials, which ones should we interview? If we are doing comparative case studies of major wars, which wars should we select? If we are interested in presidential vetoes, should we select all vetoes, all since World War II, a random sample, or only those overridden by Congress? No issue is so ubiquitous early in the design phase of a research project as the question: which cases (or more precisely, which observations) should we select for study? In qualitative research, the decision as to which observations to select is crucial for the outcome of the research and the degree to which it can produce determinate and reliable results.

As we have seen in section 4.2, random selection is not generally appropriate in small- n research. But abandoning randomness opens the door to many sources of bias. The most obvious example is when we, knowing what we want to see as the outcome of the research (the confirmation of a favorite hypothesis), subtly or not so subtly select observations on the basis of combinations of the independent and dependent variables that support the desired conclusion. Suppose we believe that American investment in third world countries is a prime cause of internal violence, and then we select a set of nations with major U.S. investments in which there has been a good deal of internal violence and another set of nations where there is neither investment nor violence. There are other observations that illustrate the other combinations (large investment and no violence, or no small investment and large violence) but they are “conveniently” left out. Most selection bias is not as blatant as this, but since selection criteria in qualitative research are often implicit and selection is often made without any self-conscious attempt to evaluate potential biases, there are many opportunities to allow bias subtly to intrude on our selection procedures.⁵

⁵ This example is a good illustration of what makes science distinctive. When we introduce this bias in order to support the conclusion we want, we are not behaving as social scientists ought to behave, but rather the way many of us behave when we are in political arguments in which we are defending a political position we cherish. We often select examples that prove our point. When we engage in research, we should try to get all

4.3.1 Selection on the Dependent Variable

Random selection with a large- n allows us to ignore the relationship between the selection criteria and other variables in our analysis. Once we move away from random selection, we should consider how the criteria used relate to each variable. That brings us to a basic and obvious rule: *selection should allow for the possibility of at least some variation on the dependent variable*. This point seems so obvious that we would think it hardly needs to be mentioned. How can we explain variations on a dependent variable if it does not vary? Unfortunately, the literature is full of work that makes just this mistake of failing to let the dependent variable vary; for example, research that tries to explain the outbreak of war with studies only of wars, the onset of revolutions with studies only of revolutions, or patterns of voter turnout with interviews only of nonvoters.⁶

We said in chapter 1 that good social scientists frequently thrive on anomalies that need to be explained. One consequence of this orientation is that investigators, particularly qualitative researchers, may select observations having a common, puzzling outcome, such as the social revolutions that occurred in France in the eighteenth century and those that occurred in France and China in the twentieth (Skocpol 1979). Such a choice of observations represents selection on the dependent variable, and therefore risks the selection bias discussed in this section. When observations are selected on the basis of a particular value of the dependent variable, nothing whatsoever can be learned about the causes of the dependent variable without taking into account other instances when the dependent variable takes on other values. For example, Theda Skocpol (1979) partially solves this problem in her research by explicitly including some limited information about “moments of revolutionary crisis” (Skocpol 1984:380) in seventeenth-century England, nineteenth-century Prussia/Germany, and nineteenth-century Japan. She views these observations as “control cases,” although they are discussed in much less detail than her principal cases. The bias induced by selecting on the dependent variable does not imply that we should never take into account values of the dependent variable when designing research. What it does mean, as we

observations if possible. If selection is required, we should attempt to get those observations which are pivotal in deciding the question of interest, not those which merely support our position.

⁶ In this section, we do not consider the possibility that a specific research project that is designed not to let the dependent variable change at all is part of a larger research program and therefore can provide useful information about causal hypotheses. We explain this point in section 4.4.

discuss below and in chapter 6, is that we must be aware of the biases introduced by such selection on the dependent variable and seek insofar as possible to correct for these biases.

There is also a milder and more common version of the problem of selection on the dependent variable. In some instances, the research design does allow variation on the dependent variable but that variation is truncated: that is, we limit our observations to less than the full range of variation on the dependent variable that exists in the real world. In these cases, something can be said about the causes of the dependent variable; but the inferences are likely to be biased since, if the explanatory variables do not take into account the selection rule, *any selection rule correlated with the dependent variable attenuates estimates of causal effects on average* (see Achen, 1986; King 1989: chapter 9). In quantitative research, this result means that numerical estimates of causal effects will be closer to zero than they really are. In qualitative research, selection bias will mean that the true causal effect is larger than the qualitative researcher is led to believe (unless of course the researcher is aware of our argument and adjusts his or her estimates accordingly). If we know selection bias exists and have no way to get around it by drawing a better sample, these results indicate that our estimate at least gives, on average, a lower bound to the true causal effect. The extent to which we underestimate the causal effect depends on the severity of the selection bias (the extent to which the selection rule is correlated with the dependent variable), about which we should have at least some idea, if not detailed evidence.

The cases of extreme selection bias—where there is by design no variation on the dependent variable—are easy to deal with: avoid them! We will not learn about causal effects from them. The modified form of selection bias, in which observations are selected in a manner related to the dependent variable, may be harder to avoid since we may not have access to all the observations we want. But fortunately the effects of this bias are not as devastating since we can learn something; our inferences might be biased but they will be so in a predictable way that we *can* compensate for. The following examples illustrate this point.

Given that we will often be forced to choose observations in a manner correlated with the dependent variable, and we therefore have selection bias, it is worthwhile to see whether we can still extract some useful information. Figure 4.1, a simple pictorial model of selection bias, shows that we can. Each dot is an observation (a person, for example). The horizontal axis is the explanatory variable (for example, number of accounting courses taken in business school). The vertical axis is the dependent variable (for example, starting salary in the first

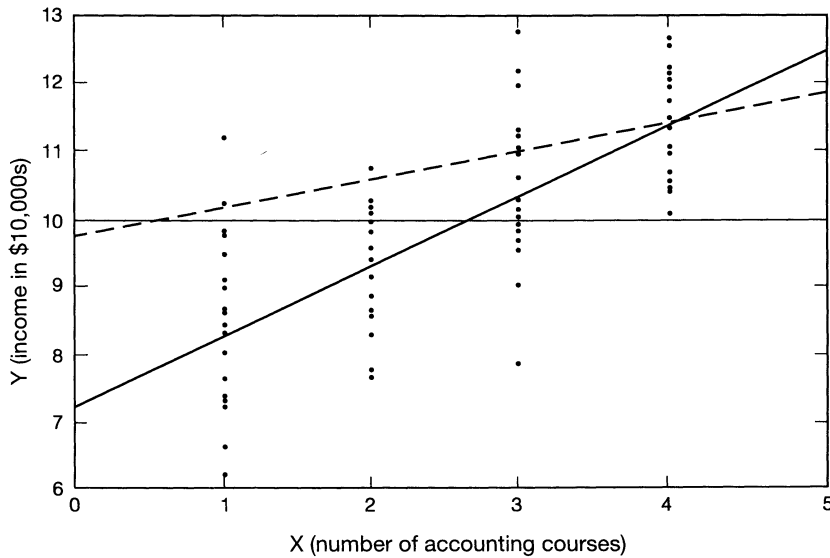


Figure 4.1 Selection Bias

full-time job, in units of \$10,000). The regression line showing the relationship between these two variables is the solid line fit to the scatter of points. Each additional accounting course is worth on average about an additional \$10,000 in starting salary. The scatter of points around this line indicates that, as usual, the regression line does not fit each student's situation perfectly. In figures like these, the *vertical* deviations between the points and the line represent the errors in predictions (given particular values of the explanatory variables) and are therefore minimized in fitting a line to the points.

Now suppose an incoming business-school student were interested in studying how he could increase his starting salary upon graduation. Not having learned about selection bias, this student decides to choose for study a sample of previous students composed only of those who did well in their first job—the ones who received jobs he would like. It may *seem* that if he wants to learn about how to earn more money it would be best to focus only on those with high earnings, but this reasoning is fallacious. For simplicity, suppose the choice included only those making at least \$100,000. This sample selection rule is portrayed in figure 4.1 by a solid horizontal line at $Y = 10$, where only the points above the line are included in this student's study. Now, instead of fitting a regression line to all the points, he fits a line (the dashed line) only to the points in his sample. Selection bias exerts its effect by decreasing this line's slope compared to that of the solid line.

As a result of the selection bias, this student would incorrectly conclude that each additional accounting course is worth only about \$5,000.

This is a specific example of the way in which we can underestimate a causal effect when we have selection on the dependent variable. Luckily, there *is* something our student can do about his problem. Suppose after this student completes business school, he gets bored with making money and goes to graduate school in one of the social sciences where he learns about selection bias. He is very busy preparing for comprehensive examinations, so he does not have the time to redo his study properly. Nevertheless, he does know that his starting salary would have increased by some amount significantly *more* than his estimate of \$5,000 for each additional accounting class. Since his selection rule was quite severe (indeed it was deterministic), he concludes that he would have made more money in business if he had taken additional accounting classes—but having decided not to maximize his income (who would enter graduate school with that in mind?)—he is thankful that he did not learn about selection bias until his values had changed.

4.3.1.1 EXAMPLES OF INVESTIGATOR-INDUCED SELECTION BIAS

The problem just described is common in qualitative research (see Geddes 1990). It can arise from a procedure as apparently innocuous as selecting cases based on available data, if data availability is related to the dependent variable. For instance, suppose we are interested in the determinants of presidential involvement in significant foreign policy decisions during recent years and that we propose to study those decisions on which information about the president's participation in meetings is available. The problem with this research design is that the selection rule (information availability) is probably correlated with relatively low levels of presidential involvement (the dependent variable) since the more secret meetings, which will not be available to us, are likely to have involved the president more fully than those whose deliberations have become public. Hence the set of observations on which information is available will overrepresent events with lower presidential involvement, thus biasing our inferences about the determinants of presidential involvement.

The reasoning used in our business-school example can help us learn about the consequences of unavoidable selection bias in qualitative research. Suppose, in the study just mentioned, we were interested in whether presidents are more involved when the events entail threats of force than when no such threats were made. Suppose also that existing evidence, based on perhaps two dozen observations, indi-

cates that such a relationship does exist, but that its magnitude is surprisingly small. To assess the degree of selection bias in this research, we would first compile a list of foreign policy situations in which the president took action or made public pronouncements, regardless of whether we had any information on decision-making processes. This list would avoid one source of selection bias that we had identified: greater secrecy with respect to decision-making involving threats of force. Our new list would not be a complete census of issues in which the president was engaged, since it would miss covert operations and those on which no actions were taken, but it would be a larger list than our original one, which required information about decision-making. We could then compare the two lists to ascertain whether (as we suspect) cases on which we had decision-making information were biased against those in which force was used or threatened. If so, we could reasonably infer that the true relationship was probably even stronger than it seemed from our original analysis.

The problem of selection bias appears often in comparative politics when researchers need to travel to particular places to study their subject matter. They often have limited options when it comes to choosing what units to study since some governments restrict access by foreign scholars. Unfortunately, the refusal to allow access may be correlated with the dependent variable in which the scholar is interested. A researcher who wanted to explain the liberalization of authoritarian regimes on the basis of the tactics used by dissident groups might produce biased results, especially if she only studied those places that allowed her to enter, since the factors that led the regime to allow her in would probably be correlated with the dependent variable, liberalization. We obviously do not advise clandestine research in inhospitable places. But we do advise self-conscious awareness of these problems and imagination in finding alternative data sources when on-site data are unavailable. Recognition of these difficulties could also lead to revision of our research designs to deal with the realities of scholarly access around the world. If no data solution is available, then we might be able to use these results on selection bias at least to learn in which direction our results will be biased—and thus perhaps provide a partial correction to the inevitable selection bias in a study like this. That is, if selection bias is unavoidable, we should analyze the problem and ascertain the direction and, if possible, the magnitude of the bias, then use this information to adjust our original estimates in the right direction.

Selection bias is such an endemic problem that it may be useful to consider some more examples. Consider a recent work by Michael Porter (1990). Porter was interested in the sources of what he called

“competitive advantage” for contemporary industries and firms. He designed a large-scale research project with ten national teams to study the subject. In selecting the ten nations for analysis, he chose, in his words, “ones that already compete successfully in a range of such industries, or, in the case of Korea and Singapore, show signs of an improving ability to do so” (Porter 1990:22). In his eagerness to explore the puzzle that interested him, Porter intentionally selected on his dependent variable, making his observed dependent variable nearly constant. As a result, any attempts by Porter, or anyone else using these data at this level of analysis, to explain variations in success among his ten countries will produce seriously biased causal effects.

But what Porter did—try to determine the circumstances and policies associated with competitive success—was somewhat related to Mill’s method of agreement. This method is not a bad first attempt at the problem, in that it enabled Porter to develop some hypotheses about the causes of competitive advantage by seeing what these nations have in common; however, his research design made it impossible to evaluate any individual causal effect.

More serious is the logical flaw in the method: without a control group of nations (that is, with his explanatory variable set to other values), he cannot determine whether the absence of the hypothesized causal variables is associated with competitive failure. Thus, he has no way of knowing whether the conditions he has associated with success are not also associated with failure. In his provocative work, Porter has presented a fascinating set of *hypotheses* based on his cases of success, but without a range of competitive successes and failures (or selection based on something other than his dependent variable) he has no way of knowing whether he is totally right, completely wrong, or somewhere in between.⁷

A striking example of selection bias is found in the foreign policy literature dealing with deterrence: that is, “the use of threats to induce the opponents to behave in desirable ways” (Achen and Snidal 1989: 151). Students of deterrence have often examined “acute crises”—that is, those that have not been deterred at an earlier stage in the process of political calculation, signalling, and action. For descriptive pur-

⁷ Porter claims to have numerous examples of countries which were not successful; however, these are introduced in his analyses by way of selectively chosen anecdotes and are not studied with similar methods as his original ten. When nonsystematically selecting supporting examples from the infinite range of supporting and nonsupporting possibilities, it is much too easy to fool ourselves into finding a relationship when none exists. We take no position on whether Porter’s hypotheses are correct and only wish to point out that the information needed to make this decision must be collected more systematically.

poses, there is much to be said for such a focus, at least initially: as in Porter's emphasis on competitive success, the observer is able to describe the most significant episodes of interest and may be enabled to formulate hypotheses about the causes of observed outcomes. But as a basis for inference (and without appropriate corrections), such a biased set of observations is seriously flawed because instances in which deterrence has worked (at earlier stages in the process) have been systematically excluded from the set of observations to be analyzed. "When the cases are then misused to estimate the success rate of deterrence, the design induces a 'selection bias' of the sort familiar from policy-evaluation research" (Achen and Snidal 1989:162).

4.3.1.2 EXAMPLES OF SELECTION BIAS INDUCED BY THE WORLD

Does choosing a census of observations, instead of a sample, enable us to avoid selection bias? We might think so since there was apparently no selection at all, but this is not always correct. For example, suppose we wish to make a descriptive inference by estimating the strength of support for the Liberal party in New York State. Our dependent variable is the percent of the vote in New York State Assembly districts cast for the candidate (or candidates) endorsed by the Liberal party. The problem here is that the party often chooses not to endorse candidates in many electoral districts. If they do not endorse candidates in districts where they feel sure that they will lose (which seems to be the case), then we will have selection bias even if we choose every district in which the Liberal party made an endorsement. *The selection process in this example is performed as part of the political process we are studying, but it can have precisely the same consequences for our study as if we caused the problem ourselves.*

This problem of bias when the selection of cases is correlated with the dependent variable is one of the most general difficulties faced by those scholars who use the historical record as the source of their evidence, and they include virtually all of us. The reason is that the processes of "history" differentially select that which remains to be observed according to a set of rules that are not always clear from the record. However, it is *essential* to discover the process by which these data are produced. Let us take an example from another field: some cultures have created sculptures in stone, others in wood. Over time, the former survive, the latter decay. This pattern led some European scholars of art to underestimate the quality and sophistication of early African art, which tended to be made of wood, because the "history" had selectively eliminated some examples of sculpture while maintaining others. The careful scholar must always evaluate the possible selection biases in the evidence that is available: what kinds of events are

likely to have been recorded; what kinds of events are likely to have been ignored?

Consider another example. Social scientists often begin with an end point that they wish to “explain”—for example, the peculiar organizational configurations of modern states. The investigator observes that at an early point in time (say, A.D. 1500) a wide variety of organizational units existed in Europe, but at a later time (say, A.D. 1900), all, or almost all, important units were national states. What the researcher should do is begin with units in 1500 and explain later organizational forms in terms of a limited number of variables. Many of the units of analysis would have disappeared in the interim, because they lost wars or were otherwise amalgamated into larger entities; others would have survived. Careful categorization could thus yield a dependent variable that would index whether the entity that became a national state is still in existence in 1900; or if not, when it disappeared.

However, what many historical researchers inadvertently do is quite different. They begin, as Charles Tilly (1975: 15) has observed, by doing *retrospective* research: selecting “a small number of West European states still existing in the nineteenth and twentieth centuries for comparison.” Unfortunately for such investigators, “England, France, and even Spain are *survivors* of a ruthless competition in which most contenders lost.” The Europe of 1500 included some five hundred more or less independent political units, the Europe of 1900 about twenty-five. The German state did not exist in 1500, or even 1800. Comparing the histories of France, Germany, Spain, Belgium, and England (or, for that matter, any other set of modern Western European countries) for illumination on the processes of state-making weights the whole inquiry toward a certain kind of outcome which was, in fact, quite rare.

Such a procedure therefore selects on the basis of one value of the dependent variable—survival in the year 1900. It will bias the investigator’s results, on average reducing the attributed effects of explanatory variables that distinguish the surviving states from their less durable counterparts. Tilly and his colleagues (1975), recognizing the selection bias problem, moved from a *retrospective* toward a *prospective* formulation of their research problem. Suppose, however, that such a huge effort had not been possible, or suppose they wished to collect the best available evidence in preparation for their larger study. They could have reanalyzed the available retrospective studies, inferring that those studies’ estimates of causal effects were in most observations biased downward. They would need to remember that, even if the criteria described above do apply exactly, any one application might overestimate or underestimate the causal effect. The best

guess of the true causal effect, based on the flawed retrospective studies, however, would be that the causal effects were underestimated at least on average—if we assume that the rules above do apply and the criteria for selection were correlated with the dependent variable.

4.3.2 Selection on an Explanatory Variable

Selecting observations for inclusion in a study according to the categories of the key causal explanatory variable causes no inference problems. The reason is that our selection procedure does not predetermine the outcome of our study, since we have not restricted the degree of possible variation in the dependent variable. By limiting the range of our key causal variable, we may limit the generality of our conclusion or the certainty with which we can legitimately hold it, but we do not introduce bias. By selecting cases on the basis of values of this variable, we can control for that variable in our case selection. Bias is not introduced even if the causal variable is correlated with the dependent variable since we have already controlled for this explanatory variable.⁸ Thus, it is possible to avoid bias while selecting on a variable that is correlated with the dependent variable, so long as we control for that variable in the analysis.

It is easy to see that selection on an explanatory variable causes no bias by referring again to figure 4.1. If we restricted this figure to exclude all the observations for which the explanatory variable equaled one, the logic of this figure would remain unchanged, and the correct line fit to the points would not change. The line would be somewhat less certain, since we now have fewer observations and less information to bear on the inference problem, but on average there would be no bias.⁹

Thus, one can avoid bias by selecting cases based on the key causal variable, but we can also achieve the same objective by selecting according to the categories of a control variable (so long as it is causally prior to the key causal variable, as all control variables should be). Experiments almost always select on the explanatory variables. Units are created when we manipulate the explanatory variables (administering a drug, for example) and watch what happens to the dependent variable (whether the patient's health improves). It would be difficult to select on the dependent variable in this case, since its value is not even

⁸ In general, selection bias occurs when selecting on the dependent variable, after taking into account (or controlling for) the explanatory variables. Since one of these explanatory variables is the method of selection, we control for it and do not introduce bias.

⁹ The inference would also be less certain if the range of values of the explanatory variables were limited through this selection. See section 6.2.

known until after the experiment. However, most experiments are far from perfect, and we can make the mistake of selecting on the dependent variable by inadvertently giving some treatments to patients based on their expected response.

For another example, if we are researching the effect of racial discrimination on black children's grades in school, it would be quite reasonable to select several schools with little discrimination and some with a lot of discrimination. Even though our selection rule will be correlated with the dependent variable (blacks get lower grades in schools with more discrimination), it will not be correlated with the dependent variable *after* taking into account the effect of the explanatory variables, since the selection rule is determined by the values of one of the explanatory variables.

We can also avoid bias by selecting on an explanatory variable that is irrelevant to our study (and has no effect on our dependent variable). For example, to study the effects of discrimination on grades, suppose someone chose all schools whose names begin with the letter "A." This, of course, is not recommended, but it would cause no bias as long as this irrelevant variable is not a proxy for some other variable that is correlated with the dependent variable.

One situation in which selection by an irrelevant variable can be very useful involves secondary analysis of existing data. For example, suppose we are interested in what makes for a successful coup d'état. Our key hypothesis is that coups are more often successful when led by a military leader rather than a civilian one. Suppose we find a study of attempted coups that selected cases based on the extent to which the country had a hierarchical bureaucracy before a coup. We could use these data even if hierarchical bureaucratization is irrelevant to our research. To be safe, however, it would be easy enough to include this variable as a control in our analysis of the effects of military versus civilian leaders. We would include this control by studying the frequency of coup success for military versus civilian leaders in countries with and then without hierarchical bureaucratization. The presence of this control will help us avoid selection bias and its causal effect will indicate some possibly relevant information about the process by which the observations were really selected.

4.3.3 *Other Types of Selection Bias*

In all of the above examples, selection bias was introduced when the units were chosen according to some rule correlated with the dependent variable or correlated with the dependent variable after the ex-

planatory variables were taken into account. With this type of selection effect, estimated causal effects are always underestimates. This is by far the most common type of selection bias in both qualitative and quantitative research. However, it is worth mentioning another type of selection bias, since its effects can be precisely the opposite and cause *overestimation* of a causal effect.

Suppose the causal effect of some variable varies over the observations. Although we have not focused on this possibility, it is a real one. In section 3.1, we defined a causal effect for a single unit and allowed the effect to differ across units. For example, suppose we were interested in the causal effect of poverty on political violence in Latin American countries. This relationship might be stronger in some countries, such as those with a recent history of political violence, than in others. In this situation, where causal effects vary over the units, a selection rule correlated with the size of the causal effect would induce bias in estimates of *average* causal effects. Hence if we conducted our study only in countries with recent histories of political violence but sought to generalize from our findings to Latin America as a whole, we would be likely to overestimate the causal effect under investigation. If we selected units with large causal effects and averaged these effects during estimation, we would get an overestimate of the average causal effect. Similarly, if we selected units with small effects, the estimate of the average causal effect would be smaller than it should be.

