CHAPTER 1

The First Rule

THERE SHOULD BE THE POSSIBILITY OF SURPRISE IN SOCIAL RESEARCH

Social research differs fundamentally from advocacy research. Advocacy research refers to research that sifts through evidence to argue a predetermined position. Lawyers engage in advocacy research in attempting to prove the guilt or innocence of defendants. Special interest groups engage in advocacy research when trying to influence the vote of lawmakers on a particular bill. The aim of advocacy research is to convince others (juries, lawmakers) to think or act in a given way. Hence advocacy research presents evidence selectively, focusing on supporting evidence and suppressing contrary or inconvenient evidence.

Social research, in contrast, does not suppress contrary or inconvenient evidence. In the social sciences we may begin with a point of view, but (unlike advocacy research) we do not want that point of view to determine the outcome of our research. Rule 1 is not intended to suggest that you need to check your idealism at the door when you embark on a social research project. Far from that—if you don't want to change the world for the better, you might not have the doggedness needed to do good social research. But rule 1 is intended to warn that you don't want to be blinded by preconceived ideas so that you fail to look for contrary evidence, or you fail to recognize contrary evidence when you do encounter it, or you recognize contrary evidence but suppress it and refuse to accept your findings for what they appear to say.

You must be ready to be surprised by your research findings. It is the uncertainty that makes social research exciting and rewarding. If you always knew the answer beforehand, what is the point of doing the research?

For first-time researchers, the surprise very often is in the finding of small effects where large effects were expected. Rarely do new researchers obtain findings that were as strong as they anticipated. Sometimes results are even opposite to expectations. Perhaps you were absolutely certain, for example, that in the United States women are more likely to approve of abortion than men are (not true: see exercises at the end of this chapter), or that younger adults tend to be happier than older adults are (which is also false: see exercises at the end of chapter 2).

Although you might at first be disappointed in your results, it is important to keep in mind that a *noneffect* is not a *nonfinding*. Any result is a finding. Finding no effect could be more interesting than finding a big effect, especially when conventional wisdom holds that there is a big effect. Often the most interesting results in social research are those that fly in the face of conventional wisdom. So you should not be disappointed when you do not find the big effect you expected, or when your results are inconsistent with "what everyone knows."

SELECTING A RESEARCH QUESTION

The first step in a research project is to decide upon a research question. A research question in the social sciences is a question about the social world that you try to answer through the analysis of empirical data. The empirical data may be data that you collected yourself, or data collected by someone else.

There are two fundamental criteria for a research question: interest and researchability. In selecting a research question, your aim is to find a question that is (1) researchable and (2) interesting to you and to others in your field.

RESEARCHABLE QUESTIONS

Some research questions in the social sciences are not researchable because they are simply unanswerable (see Lieberson 1985, chap. 1). In that respect economics, political science, psychology, and sociology are the same as astronomy, biology, chemistry, and physics. Some questions—such as the existence of God—are inherently unknowable with the scientific method. Other questions might be unanswerable as the questions are currently conceived. Other questions might be unanswerable with the knowledge and methods we currently have.

If you are a student, some questions that are answerable in principle might nonetheless be beyond your reach. The Minneapolis Domestic Violence Experiment, for example, sought to determine whether repeat incidents of domestic violence would decline if police arrested accused abusers on the spot (Sherman 1992; Sherman and Berk 1984). Typically students do not have the time, resources, or credentials required to carry out an experiment of this magnitude. The possibilities for student projects are not as restricted as one might think, however, because there are a surprising number of data sets that are available to students for analysis (the student exercises at the end of this chapter use just such a data set, the General

Social Survey). Even if you plan to collect your own survey data, you should consult such prior surveys early in your project, if only for guidance on how to word your questions.

Other research is ruled out because it is unethical. To jump ahead to an example that is central in a subsequent discussion of causality (chapter 5), we did not know, until a few decades ago, whether or not smoking causes lung cancer—some experts said it did, some said that the evidence was inconclusive. Scientifically, the best way to settle the dispute would have been to assign individuals randomly to one of two groups, a smoking group and a nonsmoking group. In terms of science, then, the question could be answered by doing an experiment. In practice, however, such an experiment would have raised serious ethical questions regarding an individual's right to smoke (or not smoke). The important point here is that ethical considerations are important—important enough to render some types of research undoable.[1]

What, then, does a researchable question look like? Researchable questions in the social sciences tend to be questions that are neither too specific (for example, about a specific individual or event) nor too grand ("Why are there wars?"). The question "Why is my uncle an alcoholic?" is not a good research question because in the final analysis it is impossible to predict the behavior of a given individual; social laws are not deterministic. (Besides, the question isn't very interesting to anyone outside your family.) It is possible to make that question researchable by generalizing it to alcoholics in general. We could ask, for example, why some people tend to be more prone to alcoholism than others are. To address that question, we first ask what characteristics distinguish alcoholics from nonalcoholics. In other words, we ask what individual or community traits *correlate* with alcoholism.

In the social sciences *what* questions generally are easier to answer than *why* questions. As Lieberson (1985, p. 219, italics removed) puts it, "Empirical data can tell us what is happening far more readily than they can tell us why it is happening." Consider alcoholism again. With appropriate data it is a straightforward matter to find the correlates of alcoholism, that is, to find the characteristics that distinguish alcoholics from others. Whether these correlates are *causes* of alcoholism is another matter (we take up the causality question under rule 5, "Compare like with

---

[1] In large part because of famous (or infamous) experiments over the past century that harmed human subjects or placed them at risk of harm, universities now maintain guidelines regarding any investigation involving human subjects. At the outset of your research you should contact your university's institutional review board (IRB) for the procedures to follow in your own research. The central principles are that participation by human subjects should be voluntary, that anonymity or confidentiality of subjects should be maintained, and that the benefits of the research should outweigh any foreseeable risks to the subjects.

like"). But that does not make the correlates uninformative. Even if we want to determine causes, generally the first step is to find correlates, since knowing the *what* gives us clues about the *why*.

Because "mere description" is sometimes devalued in the social sciences, it is important to underscore the point that "what" questions come before "why" questions. Facts come first, as Sherlock Holmes stated famously in his warning to Dr. Watson that "It is a capital mistake to theorize in advance of the facts" ("The Adventure of the Second Stain"). To punctuate the detective's point, later I describe an instance where social theorizing went astray by attempting to account for "rising global inequality" when in fact global inequality was not rising. Obviously we need to get the facts right about what there is to explain before we concoct an explanation. As the first maxim of Galileo's *Discors* states, "description first, explanation second" (see Pearl 2000, pp. 334–35).

Getting the facts right is critical, second, because accurate description sometimes is itself the end goal. Consider the problem of motion. Medieval attempts to understand motion grappled with the nature and origins of motion. Current understanding takes motion as given and attempts to describe it precisely and accurately. As Andrew Abbott (2003, p. 7) points out:

> [Isaac Newton] solved the problem of motion by simply assuming that (a) motion exists and (b) it tends to persist. By means of these assumptions (really a matter of declaring victory, as we would now put it), he was able to develop and systematize a general account of the regularities of motion in the physical world. That is, by giving up on the *why* question, he almost completely answered the *what* question.

In this instance progress was stalled until the question was switched from the unknowable (why things move) to the knowable (how they move). We have returned, then, to the central point of this section: that a research question must first of all be answerable. The second requirement is that it is interesting.

Interesting Questions

Besides finding a question that is answerable, you want a question that is interesting—to you and to others. Research involves entering into a conversation. The issue is how your research will contribute to that conversation. A research question might be interesting because of its scientific relevance: In addressing this issue, you join an ongoing discussion in the social sciences. Or a question might be interesting because it is socially important;

perhaps it bears on some important local or national social problem. Or a question might be interesting because of its timeliness; a question relating to Americans' attitudes about gun control might be particularly interesting, for example, when important legislation is pending on the subject.

"The heart of good work," Abbott (2003, p. xi) writes, "is a puzzle and an idea." An interesting research question encompasses both. The puzzle is the issue about the social world that bears investigation. The idea is the new twist that you bring to the investigation.

How do you find an interesting research question? Often personal experiences, or the experiences of others we know, provoke our curiosity about some aspect of human behavior. You can also obtain ideas by reading the research of others. New findings often lead to new questions, and research reports often conclude by noting further research that is needed, or by suggesting directions that subsequent research might take.

Unfortunately, there is no foolproof recipe for cooking up an interesting research question. In that light, it is important to keep in mind that the ideas below are intended to stimulate your thinking, not to provide a roadmap to discovery. No one knows where lightning will strike, but it is still a good idea to avoid sitting on a high metal roof in a thunderstorm. Inspiration is the same way: We cannot be certain where it will strike, but we know that the probabilities are greater in some places than in others.

If you have trouble thinking of a good question for a research project, most likely the problem is either that you have an idea that you like but others don't, or you cannot come up with an idea in the first place.

Suppose you have an idea for a project that enthuses you but no one else. One reaction—perhaps most common among students doing their first research project involving actual data analysis—is to ignore what others, such as your professors, think. That is generally not a good idea. The interest criterion for a research question dictates that you find an issue that is interesting to you and to others in your field. The "others in your field" requirement is important because the "so what?" question is the first challenge you will face when writing up the results of your research. Early in the report you must be able to explain, in a few sentences, why your results are of interest to your readers. Perhaps the question you address relates to an ongoing discussion in your field. Perhaps it extends prior research in new directions, or replicates findings using a new population. Or perhaps it sheds light on an empirical puzzle. The point is that you need some rationale for doing your research. The fact that you cannot get anyone else interested in your research is a telltale sign that you need to rethink the project. Rethinking the project does not necessarily mean that you need to abandon your idea. But it does mean that you need to rework your question to make it interesting to others.

The issue, then, is how to revise your research question to make it more interesting to others. (I assume here that you have done the necessary preliminary review of other research on the topic, so you have a basic idea of what others have found before you.[2]) First you should discuss the project with others to find out why it is uninteresting to them. Most likely they will say that your question is uninteresting either because we already know the answer, or because we would not care about the answer even if we knew it. That is:

- **The no-surprises objection**: We won't learn anything from the research because the answer either is already well documented by prior research or is preordained by the research method. In other words, the research is gratuitous, since we know the answer before we do the research.
- **The "so what" objection**: The answer to the research question either has no relevance for social science theory or for everyday life, or matters to such a small handful of people that the question is trivial.

"Why is my uncle an alcoholic?" is a good example of a trivial research question. No matter how important the answer might be to you, the question is trivial to social scientists because the answer matters to such a small handful of people. To make the question interesting we must generalize it. Instead of asking why your uncle is an alcoholic, we ask what general traits distinguish alcoholics from others. So the triviality problem generally is solved by casting a wider net in the research question.

Irrelevance is also solved by casting a wider net. In truth, of course, few questions are completely irrelevant; relevance is a matter of degree. As editor of the *American Sociological Review* I found that one of the most frequent complaints of reviewers was "Why should I care about the results of this paper?" As social scientists we care about research when the results speak to ongoing conversations and debates in our field. The more such connections we can make, the more interesting our results tend to be. So it is important to show how our research question links to theories and other empirical work.

The other objection (no-surprises) is consequential because, as rule 1 states, surprise is a hallmark of social research. It is the possibility of

[2] In early stages of your literature search you should look particularly for review articles. A good review article will provide (1) a theoretical grid or template for thinking about the important issues in the field, (2) an overview of key findings and key unresolved issues, and (3) a description of the most influential and important studies in the area. The *Journal of Economic Perspectives* and the *Journal of Economic Literature* are valuable sources of summary articles of the latest findings in particular areas of economics. In the other social sciences review articles are available in the *Annual Review* series (e.g., *Annual Review of Sociology*). When beginning a new research project it is a good idea to scan the table of contents of the *Annual Review* to see if recent summary articles are available, and go from there.

surprise that distinguishes social research from advocacy research. So if your social research project is uninteresting because it lacks the element of surprise, you need to make the case that the answer is *not* known beforehand. Sometimes results can be known ahead of time (making the research uninteresting) because the results are built in by the way concepts are measured. So if others say your project is uninteresting, you first need to make sure that your method is not to some degree preordaining your result. In examining the effect of marital satisfaction on overall life satisfaction, for example, it is important to find independent measures of the two concepts, so that the measures used do not guarantee a positive relationship between the two.

Alternatively, perhaps the possibility of surprise has been eliminated by prior research; you've been "scooped." Lest you be too hasty in that judgment, however, it is important to distinguish *well-established research* from *well-established findings*. Lots of research does not necessarily mean that researchers have reached a consensus. Indeed, because controversy itself can stimulate research, an extensive research literature might signal the opposite. Because research in the social sciences sometimes yields mixed, inconclusive, or controversial findings, there are issues of import where research is well established but findings are not. If your research falls into that category, then it does not run afoul of the no-surprises objection, since prior findings do not tell you what you will find. In the happiest of circumstances, you will devise a critical test that reaches a definitive conclusion and shows why prior research has been inconclusive. So there is great potential for a significant research contribution where there is much research but little in the way of established findings.

That situation is quite different from the situation where there is general consensus on research findings. Unless prior research is seriously flawed, well-established findings from prior research rule out the possibility of surprise. So if you believe, for example, that prior research has amply demonstrated that married people in the United States tend to be happier than unmarried people, there is little point in another study of the association of marital status and happiness unless the new study extends that finding in some new direction (examples below). Note the critical qualifier here: *if you believe prior research has amply demonstrated*. Some of the most interesting and significant research in science has overturned the findings of prior research. There is nothing wrong with challenging well-established findings in the social sciences *if you can show that there is reason to doubt those findings*.

In short, if you want to investigate a topic with (apparently) well-established findings, you have two options: *challenge the findings* or *clarify/extend the findings*. Challenging prior research involves both deductive and inductive work: First you demonstrate the logical or methodological

flaw in prior research (deduction), then you show, empirically, that the flaw matters (induction).

The other option is to clarify or extend established findings. The *clarification* of findings is deductive (or largely so), involving further conceptualization and theorizing. The *extension* of findings is inductive, involving further data analysis. I consider each of these options in turn.

Challenging Prior Research

In challenging the findings of prior research, the burden of proof rests with you to identify some shortcoming or flaw that is serious enough to raise questions about the reliability of earlier results. Personal anecdotes are not enough in this regard. For example, with regard to the established finding that married people tend to be happier, it won't do to say "I'm not married and I am happier than most married people I know" or "I know lots of unmarried people who are happy and lots of married people who aren't." If you believe prior results are off the mark, you need to focus on the relevant research itself and identify precisely where it has gone astray. Perhaps key concepts are not measured adequately, or the statistical methods are misapplied, or there are problems with the sample or the data.

Box 1.1 gives an example of a presumably well-established social science finding that was effectively challenged by new research. Early accounts suggested that global income inequality shot up dramatically in the last decades of the twentieth century. Using more appropriate data and methods to analyze the trend over that period, however, recent research now finds that global income inequality did not grow at all over that period, and most likely declined somewhat.

The example in box 1.1 is not typical. Not all research is easily challenged. In the first place, you must find something clearly dubious about earlier research, and explain how your suspicions could account for the observed findings. You might, for example, question the finding about marriage and happiness on the grounds that the finding is based on self-reported happiness, and you have good reason to believe that self-reported happiness is an unreliable measure. Your argument might be that, because American culture places a premium on being married, the married are more likely to say they are happy whether they are or not, and the unmarried are more likely to say they are unhappy whether they are or not, and it is that bias in the measurement of happiness that accounts for the association between marriage and happiness.

The next step is to devise a test to show that your critique matters empirically. In the example of marriage and happiness, you would hypothesize that the association between marriage and happiness disappears when happiness is measured in (presumably) more reliable ways, such as how

others rate you on level of happiness. Very possibly you will be wrong—you may find that marriage is associated with happiness no matter how you measure happiness. Your study nonetheless contributes by showing that marriage is related to happiness even when we measure happiness in different ways. In other words, when you use a new approach to challenge an old finding, you add to knowledge whether you are right or

---

Box 1.1
Case Study: Rethinking the Findings of Research on Global Income Inequality

Contrary to some earlier research, recent studies of global income inequality find that, when China and India are given their due weight, global income inequality is not rising (Bourguignon and Morrisson 2002; Firebaugh 2003; Goesling 2001; Melchior and Telle 2001; Sala-i-Martin 2006; Schulz 1998). (To be clear, there is massive global income inequality in the world today—but it is not growing rapidly, as once feared.) The "yes it is/no it isn't" story of rising global income inequality provides a good example of the importance of knowing what it is that you are trying to explain before you try to explain it. In a number of popular articles and books, inequality has been linked to economic globalization, one of the hottest topics in social science, to fashion a growing literature on globalization and the "explosion" of global income inequality. But the argument that globalization is the cause of rising global income inequality is wrongheaded—it tries to explain a "fact" that isn't.

It is instructive to see where the earlier research went wrong. Income inequality refers to the disproportionate distribution of income across individuals or households. Social scientists have devised ways to measure disproportionality in income (Allison 1978). By applying these measures to income data over time we can determine whether income inequality is rising or falling for a country—or even the whole world.

Global income inequality refers to the disproportionate distribution of income across all the world's citizens, where each citizen is given equal weight. With regard to the discussion here, the important point to note is that the claim of rapidly growing global income inequality in earlier research was subject to challenge because it rested on a series of dubious assumptions and incomplete analyses. These included:

---

wrong. If you are right, your results pose a challenge to conventional findings. If you are wrong, your results extend the conventional findings, albeit that was not your intention.

Probably the most common outcome is that challenges serve to refine rather than dislodge established research findings. Of course, the refinement

(continued from previous page)

- *Extrapolation from results for the tails of the income distribution.* Some earlier studies concluded that global income inequality is rising rapidly on the basis of a comparison of income trends for the very richest and very poorest countries. But the vast majority of the world's people live somewhere else. We cannot reach defensible conclusions about the inequality trend for all the world's population by examining income trends only for those who live in the very richest and very poorest regions.
- *Unequal weighting of individuals.* Early studies often weighted all countries equally. However, because global income inequality refers to the unequal distribution of income across *individuals*—and many more individuals live in poor countries (e.g., China) where incomes are growing rapidly than in poor countries where incomes are growing slowly (Firebaugh 2003)—the failure to weight countries by their population produced misleading results in those studies.
- *Extrapolation from historical trends.* Global income inequality rose rapidly over the nineteenth century and the first half of the twentieth (Bourguignon and Morrisson 2002), as the West took off economically and Asia and Africa lagged behind. Global income inequality rose rapidly because the richest regions of the world were the world's growth leaders and the poorest regions tended to grow slowly, if at all. In recent decades, however, the world has changed. Now many of the fastest-growing economies are in poorer regions of the world (Garrett 2004)—a point not fully taken into account in some early studies. Because income inequality is now declining across countries, the longstanding growth in global income inequality has halted. (At the same time, income inequality has increased *within* many countries. Declining inequality across countries and rising inequality within them implies a "new geography of global income inequality" in which countries are receding in importance as economic units: see Firebaugh [2003].)

of prior research could serve as your starting point. In that case you do not set out to challenge prior research, but to clarify or extend prior research. Indeed, unless the flaws in prior research are obvious, important, and easily correctable, clarification and extension are the most promising strategies for avoiding the no-surprises objection, especially for new researchers. In the next two sections I describe the principal ways that you might clarify or extend prior research.

Clarifying Concepts in Prior Research

You might discover that the problem with prior research is not that the findings are wrong, but that the implications are unclear because of conceptual fuzziness in the model or the variables. Eliminating the fuzziness might not lead to new results, but it could lead to new understanding about what the results mean. In this case the element of surprise in our research involves the implications of our findings, not necessarily the findings themselves.

It is especially important to be clear about the meaning of our dependent variable, the variable we are trying to explain. Richard Felson (2002, pp. 209–10) gives examples from theory and research on crime. As Felson notes, the first step in explaining crime is to determine what sorts of behavior we are trying to explain. Some types of crime involve deviant acts (for example, illegal drug use), whereas other types involve aggression as well as deviance (for example, assault). Appropriate explanations of illegal drug use may be very different from explanations of aggressive behavior. To test various theories of crime, then, Felson suggests the use of what he calls "discriminant prediction" (for example, theories of aggression should better predict assault than illegal drug use, whereas theories of deviance should predict both assault and illegal drug use).

Extending Prior Research

Suppose you have no good reason to doubt the accuracy or conceptual framework of prior research bearing on your research question. In that case you should use prior findings as the point of departure for your own research project. To make the research interesting—to permit surprises—one possibility is to extend that research in new directions. Here are some general ideas along that line.

- *Focus on a key subpopulation.* The idea here is to see whether or not earlier research findings are reproducible for some strategic subset of the population. Suppose prior research shows that college seniors tend to be more liberal politically than freshmen are. You attend a

parochial school, and in your review of the literature you find that no one has examined that issue separately for parochial schools, which (you suspect) tend to be more conservative than other colleges. In that case you might design a survey for freshmen and seniors in your school to see if you find the same result for your college (better yet, administer your survey in a broad range of parochial colleges in addition to your own). Because no one has studied this issue for parochial schools, you can be surprised by the results. The results of your study likely would be of interest to a number of people, including administrators and others in parochial schools, educational sociologists, and those who study how political attitudes are formed. (A word of warning here: subpopulations such as "my friends" or "my sorority sisters" are too narrow. You might begin by pretesting your survey on your friends, but you do not want to end there. Not many people are interested in the political attitudes of your friends or those in your sorority. But a number of people are interested in the political attitudes of students in parochial colleges, a much more substantial subset of the general college population.)

- *Use a new population.* In this case the issue is whether or not earlier research findings apply to some new population. You might want to know whether there is an association between marriage and happiness in the case of arranged marriages. To find out, you could investigate the effect of marriage on happiness in societies where arranged marriages are more common than they are in the United States.
- *Use a different time period.* Here the issue is whether or not prior research findings apply to a different era. In some studies that new time period might be historical: Was there a positive association between marriage and happiness in the nineteenth-century United States, or has that association come about more recently? Historical data to address such questions often are lacking, however, so most replication studies of this type are updating studies—studies that seek to update earlier research with more recent data. For example, studies of racial attitudes in the United States have consistently found that white southerners are more likely than whites outside the South to disapprove of marriage between blacks and whites, to disapprove of racial integration of schools, and so on. Many of those studies were done in the 1970s and 1980s (for example, Middleton 1976; Tuch 1987), however, so periodically new studies appear (for example, J. S. Carter 2005) to determine how quickly regional differences are disappearing (if at all).
- *Use a new population or subpopulation to test a specific hypothesis.* In this case you choose a population or subpopulation that provides a test of some hypothesis. For example, you might hypothesize that the positive relationship between marriage and happiness is based on

companionship, so you expect little or no effect of marriage on happiness where long periods of separation are common. You might test this hypothesis by estimating the effect of marriage on happiness among those in the navy (where long periods of spousal separation are commonplace) and comparing that effect with the marriage effect in branches of the military where separation is not the norm. Note that in this instance a *noneffect* (no association between marriage and happiness among those in the navy) provides support for your companionship explanation of the general marriage-happiness relationship.

In short, one possibility for well-established research results is to see whether or not old results apply in new contexts (strategic subpopulations, new populations, or different time periods).

Another possibility is to extend or refine the *causal explanations* of the observed empirical patterns found in prior research. The social sciences are pregnant with claims about what causes what. The evidence we have for these causal claims, however, is often very weak. A critical task in the social sciences is more and deeper research on causal relations. There are several general strategies for doing so. The major possibilities are discussed below.

CAUSAL CHAINS: INTRODUCE MEDIATING VARIABLES TO ACCOUNT
   FOR RESULTS OF PRIOR STUDIES

The idea here is to develop theories that identify the mechanisms or pathways by which some independent (explanatory, exogenous) variable X affects some dependent (outcome, endogenous) variable Y. One way to account for the effect of X on Y is to find some set of variables Q that mediate the effect of X on Y, that is, $X \rightarrow Q \rightarrow Y$. This strategy is often employed in social science to account for the effects of ascribed traits such as gender or race on outcome variables such as earnings. It is well known, for example, that women tend to earn less than men in the United States. Explanations of this gender gap in earnings typically focus on mediating or intervening variables Q in this chain: gender $\rightarrow Q \rightarrow$ earnings. In this chain Q might include things like interrupted careers (women are more likely to have interrupted work careers, and that's one reason for gender differences in earnings) and hours worked (men are more likely than women to work full-time). One way to make a research contribution, then, is to find important mediating variables that have been missed or undervalued in prior research.

The mediating variables strategy works best when reverse causation can be ruled out. That is why the strategy is often popular for explaining the effects of ascribed traits such as gender and race. In the illustration

above, for example, it is clear that gender affects hours worked, not vice versa.

Sometimes causal direction is not so obvious. In the case of marriage and happiness, it is not clear which comes first, marriage or happiness (see reverse causation below). In other words, there could be a kind of selection effect in which those who are selected into marriage tend to be happier than those who are not selected into marriage. Even if marriage is not selective with respect to happiness per se, marriage could be selective with respect to possible *correlates of happiness*, such as physical attractiveness, or *causes of happiness*, such as physical health. If in fact the beautiful and the healthy are more likely to marry *and* more likely to be happy, then at least some of the association of marriage with happiness is due to the association of marriage with attractiveness and health. Thus marriage could be associated with happiness even if marriage has no *causal* effect on happiness.

This observation—that *nonmediating* variables could account for the empirical association between marriage and happiness—leads to another possibility for extending prior research.

INTRODUCE NONMEDIATING VARIABLES TO ACCOUNT
   FOR RESULTS OF PRIOR STUDIES

The idea here is the same as in the mediating variables case: We want to develop theories that explain why we observe an association between X and Y. In the case of nonmediating variables, though, the mechanisms are not causal. Sometimes the term *spurious association* is used to describe associations between X and Y that can be accounted for by their associations with other variables, where those other variables are not causal mechanisms linking X and Y. (For example, X and Y might be associated because some third variable Z causes both of them. Or X and Y might be associated because X is associated with other variables that are causes of Y.) The term *spurious* here means that there is *no causal link* from X to Y. It does not mean that the association of X and Y itself is a mirage. When we say, for example, that the association of marriage and happiness is spurious, we do not mean that, on second thought, married people aren't happier after all. What we mean is that getting married typically won't make you happier. (If the distinction is unclear, it should become clearer with the discussion of causality in rule 5, "Compare like with like.")

Often an association between X and Y is partly causal and partly not, so some of the so-called *control variables* that are added to a model to account for an association of interest are mediating and some are not. Whatever the situation—whether the association of X and Y is due to mediating variables or nonmediating variables or a combination of the two—it is nearly impossible to overstate the importance of careful conceptualization

and theorizing in determining what control variables to add. To assist in this theorizing, you should draw a path diagram that depicts your model (chapter 5 provides examples). In a path diagram, variables are nodes, and arrows connect the variables. A straight arrow from X to Y indicates that X causes Y, and a curved double-headed arrow between X and Y indicates that X and Y are associated, with no causal direction assumed. If there are no reciprocal effects among the variables, path analysis methods (Alwin and Hauser 1975) can be used to estimate the model. The important point here, however, is to draw the model so that you can see at a glance how all the variables are interconnected.[3]

Finally, it is important to stress that you should not add control variables willy-nilly to a model to find out if that affects your results. You need to justify the variables theoretically before you add them. More control variables are not always better than fewer control variables, as Lieberson (1985, chap. 2) emphasizes in his discussion of the dangers of what he calls "counterproductive controls" and "pseudocontrols." Careful theorizing is the best protection against such dangers.

INTRODUCE MODERATING VARIABLES TO ACCOUNT FOR,
    AND QUALIFY, RESULTS OF PRIOR STUDIES

A *moderating variable* is a special kind of nonmediating variable that conditions the effect of one variable on another. For example, if the effect of education on income is greater for whites than it is for nonwhites, then we say that race *moderates* or *conditions* the effect of education on income. The effect itself is called an *interaction effect* (Jaccard and Turrisi 2003).

By introducing moderating variables, Scott Myers turns a common observation—that religious parents tend to have religious children—into an interesting research project. After noting that "Religiosity, like class, is inherited," Myers (1996, p. 858) theorizes that there are factors that would "condition the ability of parents to transmit their religiosity." In other words, he theorizes that there are interaction effects. He finds that the transmission of religious practices (such as praying, reading the Bible, and attending church) is stronger when parents are happily married and do not divorce, when parents agree on religious beliefs, when parents show affection toward their children and are moderately strict in rearing their children, and when the husband works in the paid labor force and the wife is a housewife. In short, Myers finds a number of interesting interaction effects.

[3] Path diagrams also help you to see which variables are exogenous (determined outside the model) and which are endogenous (determined by other variables in the model). It is important to find exogenous variables, since you cannot estimate a model when all variables are endogenous.

We encountered the concept of interaction effects earlier, most notably in our hypothesis that the effect of marriage on happiness will be reduced in navy marriages (as opposed to marriages among those in other branches of the military). In fact, if we think again about comparing the findings of new populations with findings for other populations, we see that such comparisons are best carried out in a moderator-variable framework, that is, by combining data for the old and new populations and using interaction terms. In the case of comparing students in parochial colleges with students in other colleges, for example, we could append the data collected on parochial schools to existing data on other colleges and employ "type of school" (parochial/other) as the moderating variable. (Merging the parochial school data with the earlier data should not be a problem if the replication was done properly.) Or, to determine whether the effect of marriage on happiness is weaker in the case of navy marriages than for marriages in general, we could combine data from two samples—a sample of those in the navy, and a sample of those not in the navy—and use "in navy" (yes/no) as the moderating variable.

CHECK FOR REVERSE CAUSATION

The extension of prior research might also turn on the issue of *which variables are dependent* (called the *endogeneity problem*). You might concede, for example, that the married in fact tend to be happier (no matter how happiness is measured), but not because marriage leads to happiness. As noted earlier, the association could reflect the fact that happier people are more likely to attract a mate, so happiness increases the likelihood of marriage instead of the other way around. If prior research interprets the marriage-happiness association as reflecting only the effect of marriage on happiness, you can make a contribution by designing research to separate the effect of marriage on happiness (marriage → happiness) from the effect of happiness on marriage (happiness → marriage). If the conventional interpretation is that marriage causes happiness and you show that happiness causes marriage instead, then you have found an example of reverse causation: There is a causal effect, but prior research has the arrow going in the wrong direction.

BROADEN THE RESEARCH QUESTION BY EXPANDING IT ACROSS SUBFIELDS

The idea here is to reframe old issues by synthesizing perspectives, concepts, or methods from more than one subfield. Synthesizing studies are hard to do, but successful studies can have high payoff in terms of new insights. Consider again the issue of the transmission of religiosity across generations. Social class is also transmitted across generations, and there is a huge sociological literature on class inheritance. It is

useful to think about how the study of religious inheritance could be linked to the study of class inheritance, which has been studied in much greater depth. Are the processes of class inheritance and religious inheritance similar? Are parents in the United States more effective in transmitting their class standing or their religious beliefs? To my knowledge no one has tackled that question (I suppose because it is difficult to find a common metric for comparing the two types of inheritance). The attempt to span the two fields could have high payoff, however, in terms of both substance and method. We could generalize the issue more—and make it even more interesting—by asking what overarching principles govern the transmission of beliefs, attitudes, values, and statuses across generations. For example, are parents generally more successful in transmitting their beliefs and values (political, religious, and so on) or their material statuses (such as their social class) to the next generation? In synthesizing studies of this sort, prior findings for particular fields of inquiry become the touchstones for more general theories and research about human behavior.

In short, the answer to the question of concern here—What if your research idea is not interesting because it has been thoroughly researched before?—is: Don't give up on your idea too easily. You might be able to update previous findings with more recent data, or extend findings back in time with historical data. Or you might try to replicate prior research by focusing on a strategic subpopulation, or by using a new population altogether.

Alternatively, you can use the same population as in prior research, and look for ways to extend that research by accounting for well-documented associations in terms of either mediating variables, nonmediating variables, moderating variables, or reverse causation (or some combination of the above). Your project can be especially interesting if it bears on *empirical regularities* in societies—fundamental associations that are well documented and appear to be generally true of most societies (for example, the inheritance of social class across generations). In general, the more tightly you can link your research to key empirical regularities, the more interesting it is for social scientists.

In offering these suggestions about how to build upon the foundation of prior research, I am assuming that there is prior research on the issue you want to study. Of course you do not need to worry about prior research if there is no prior research. So it is tempting to try to think of a novel research idea, something that no one else has investigated before. That often poses a dilemma. Recall that interest and researchability are the key requirements for a research topic. Yet if we find a topic that is

interesting and researchable, very likely the issue has been studied a number of times, so there is extensive research on it already. It is generally difficult, then, to find a feasible research question that is important yet has not been studied.

That dilemma probably explains why researchers-in-training sometimes struggle to find a research topic. They try in vain to find an interesting question that no one has investigated before. The examples above are intended to make the point that the best social research most often is research that brings fresh perspectives and new insights to old and continuing areas of concern.

SELECTING A SAMPLE

After determining *what* you want to study, you need to determine *whom* to study. First you need to determine what target population your results are meant to describe (see chapter 4). Are you interested in drawing conclusions about all college students worldwide? (That could be overly ambitious, as obtaining a sample of college students worldwide would be difficult.) Or only American college students? Or only American college students in parochial schools? To know whom to sample, you need to identify your target population carefully and precisely.

After determining your target population, you need to determine a strategy for selecting a subset of individuals—called a *sample*—from that population. Because you attempt to draw conclusions about the entire population on the basis of results from a subset of that population, it is critical that your sample be representative of your population (or the sub-populations that you want to compare: see the third principle of sampling, below). The most straightforward way to obtain a representative sample, conceptually, is to think about selecting individuals completely at random from the population, so each individual in the target population has an equal probability of being selected for the sample. A sample selected in this way is called a *simple random sample*. In practice, though, a simple random sample usually is hard to obtain, and other sampling methods have been devised for obtaining samples that are representative, or virtually so, for target populations. Full descriptions of these methods are available elsewhere (Kalton 1983; Kish 1965), and I see no need to cover the same ground here. What I want to do instead is emphasize some overarching principles.

> *First principle of sampling*: You don't need to eat the whole ox to know that it is tough.

Samuel Johnson is often credited (probably incorrectly) with this colorful way to express the idea—later applied to the sampling of human populations—that characteristics of the whole can be inferred from characteristics of the part. All sampling is based on this principle.

The same laws of probability that underlie the first principle of *sampling* also lead directly to the first principle of *sample size*.

> *First principle of sample size*: For practical purposes, very large populations do not require larger samples than smaller populations do.

Suppose we want to know, with a certain degree of precision, which of two candidates will carry the states of California and Wyoming in a presidential election. If the race is equally competitive in both states, we do not need a larger sample for California than we do for the state of Wyoming, even though California has seventy times more people than Wyoming does.[4] In fact, if the race is closer in Wyoming than it is in California, then we would need a larger sample to predict the winner in Wyoming.

The first principle of sample size implies an important corollary.

> *Sample size corollary:* We can make confident generalizations about a large population from a sample containing only a tiny fraction of that population.

Although students generally have little trouble grasping the first principle of sampling, they often are surprised by the first principle of sample size and its corollary. It seems to defy common sense that anyone could determine what millions of Americans are thinking by talking to merely 1,500 of them. Yet that is exactly what pollsters try to do.

George C. Wallace, four-term governor of Alabama and unsuccessful third-party candidate for president in 1968, provides the most memorable attacks on the reliability of political polls and the logic of polling.

---

[4] Note that the principle states *for practical purposes*. Strictly speaking, all else equal, we would need a slightly larger sample in California, due to the finite population correction $\sqrt{(N-n)/(N-1)}$, where n is sample size and N is size of the population. (This correction is needed because the standard error approaches zero as n approaches N.) In practice n is usually so small relative to N that the correction is trivial. According to U.S. Census Bureau estimates for 2003, California had 35.48 million people and Wyoming had 0.50 million people. Thus for a sample size of 1,000, the finite population correction factor is 0.999986 for California and 0.998999 for Wyoming. In other words, because a sample of 1,000 constitutes a larger fraction of the population in Wyoming than in California, the standard error shrinks more in Wyoming (by a factor of 0.998999) than in California, where the standard error is reduced only by a factor of 0.999986. The practical implication is that, all else equal, we would need a sample size of 1,001 for California to match the precision of the results for a sample of 1,000 in Wyoming.

Wallace's populist "Stand up for America" campaign resonated with large segments of the U.S. public,[5] and polls showed surprising support, particularly in the South, for his third-party candidacy. However, when polls showed his support slipping in the final weeks of the 1968 election, Wallace attacked the polls themselves. In an article "Wallace Assails 'Lying' Election Polls," the *New York Times* (October 27, 1968) reports:

> George C. Wallace delivered a blistering attack on national political polls today in response to one that indicates support for his Presidential campaign has slipped substantially in the last week. Informed of the results of the latest Gallup Poll, which shows that the American Independent party candidate dropped 5 percentage points, he said: "They lie when they poll. They are trying to rig an election. Eastern money runs everything. They are going to be pointed out the liars that they are."

Two days later Wallace dismissed political polls as "comic strips" and charged that the polls represent "a deliberate and desperate attempt on the part of the other two parties to deceive the American people" ("Wallace, Irritated by Polls, Insists He Is Doing Well," *Washington Post*, October 29, 1968, p. A7). In a campaign rally in Hannibal, Missouri, on October 28, 1968, he posed this famous question, as described the next day in the *Washington Post*: " 'The Gallup poll and the Harris poll—they talked to 1600 people in the country and say they can decide how the election's coming out,' the former Alabama Governor sniffed. 'Well, how many of you have been talked to by the Harris poll and the Gallup poll?' he demanded. 'Not a one of you, not a one of you.' "

Governor Wallace's views notwithstanding, reputable preelection polls of likely voters have been fairly accurate in forecasting the results of recent elections, even though the polls typically are based on a sample of only a dozen or so respondents per *million* voters. On the eve of a presidential election, for example, most national polls rely on a sample of 1,500 or so likely voters. Over 122 million Americans voted in the 2004 election—so a sample size of 1,500 works out to a ratio of one person sampled for every 81,000 voters, or 0.0012 percent of the voters.

How then are we to account for the fact that we can use 0.0012 percent of the population to predict so accurately the behavior of the remaining

---

[5] Wallace's campaign brochure describes him as "the undisputed leader in the fight for personal and property rights, and against excessive taxation and the takeover of personal rights by the 'great society.' " Regarding the Vietnam War, Wallace is quoted as follows in the brochure: "These few people today who are out advocating sedition and raising money and clothes and supplies for the Viet Cong—these college professors who are making speeches advocating victory for the Viet Cong Communists—I would deal with these people as they ought to be dealt with, as traitors." (http://www.4president.org/brochures/wallace1968brochure)

99.9988 percent? This principle is demystified somewhat by the second principle of sampling.

> *Second principle of sampling: How* cases are selected for a sample is more important than *how many* cases are selected.

In other words, the *representativeness* of the sample is more important than the size of the sample. Of course I do not mean to suggest that a sample of two or three is all right, so long as those people are selected properly. But I do mean to suggest that a sample of only a few hundred can give a surprisingly accurate picture of a population of many millions, so long as the individuals in the sample are representative of the individuals in the target population. Obviously a larger sample is better than a smaller sample, other things equal—but the superiority of larger samples over smaller ones is itself due to the fact that (for a given sampling method) larger samples are more likely to be representative. Consider the extreme case of a single individual. A sample of one cannot be very representative of the entire population, unless individuals in the population are all alike. At the other extreme, a census is clearly representative of the population, since it contains everyone in the population.

It is important to understand two points here: first, that a representative sample of a few hundred people can be remarkably accurate; second, that a nonrepresentative ("biased") sample of a few million people can be remarkably inaccurate. Failure to appreciate fully the second point has led to some infamous blunders in the polling business. In the United States, "the poll that changed polling" occurred in the 1936 presidential election between Democrat Franklin D. Roosevelt and Republican Alfred M. Landon. In its October 31, 1936, issue, the venerable magazine *Literary Digest* predicted—on the basis of a sample of *over two million*—that Landon would win the election in a landslide. Three days later, Roosevelt won in a landslide. The error by the *Literary Digest* was caused by their sampling method, which oversampled Republicans and those favoring change in the government (Roosevelt was the incumbent). On the basis of a more representative sample of just 5,000 people, a little-known pollster named George Gallup correctly predicted Roosevelt's landslide victory. Not long after the 1936 election, the *Literary Digest* went out of business. The polling methods perfected by Gallup are still in use today.[6]

Because scientific polling has come a long way since the 1936 presidential election, it is hard for individual researchers to match the general

---

[6] Controversies over the results of reputable polls (such as Roper, Harrris, or Gallup polls) rarely center on sampling issues any more. Most of the focus today is on validity issues—on what respondents really mean by their responses (see Sudman, Bradburn, and Schwarz 1995; Tourangeau, Rips, and Rasinski 2000).

quality of data collected by professional firms. Most social scientists, then, rely on survey firms for the collection of the survey data they use. That is one reason for the popularity of *secondary analysis* in the social sciences, that is, the analysis of data collected by someone else. Of course, with secondary analysis of national data you cannot tailor the data to very specific or local populations—you would not, for example, be able to study the political attitudes of students in your own college with national survey data. National survey data on the attitudes of college students do, however, provide a benchmark for the attitudes of students in your own college. In short, secondary analysis of national survey or census data very often is called for even in studies where researchers collect their own data, since the national data provide a point of comparison for the more local or specialized data.

Large survey data sets have become increasingly accessible and user-friendly over recent decades. Many of these data sets have been underwritten by government money, and researchers who use public funds to collect data generally are required to place their data in the public domain for others to use. Major social science data sets are located in several data archives in the United States and other countries. Probably the best-known U.S. social science data archive is housed at the University of Michigan, at the Inter-University Consortium for Political and Social Research (ICPSR), web site http://www.icpsr.umich.edu. Undergraduate and graduate students typically have access to the national data sets most commonly used in social research.

> *Third principle of sampling*: Collect a sample that permits powerful contrasts for the effects of interest.

If you are interested in the effects of cohabitation, for example, then make sure you have enough cohabiters in your sample to permit the comparison of cohabiters with others. If you are collecting your own data, you will want to *oversample* cohabiters. You can oversample cohabiters by using a method called *stratified random sampling*: First stratify on the basis of union type (cohabiting versus married); then randomly sample within each group. To oversample cohabiters, simply select a larger fraction of cohabiters than marrieds.

As this example suggests, simple random sampling of the overall population is not always the best method. Sometimes it is better to divide the population into homogeneous groups and oversample the minority groups. If there are *G* groups, we can think of *G* samples, one for each group. So long as you randomly sample within each group, each sample is representative of that group, and there is no problem in making comparisons *across* groups. But the *G* samples merged are *not* representative of the overall population. Suppose, for example, that a preelection poll heavily

oversamples African Americans. To predict results for each candidate, you would need to downweight the African American sample to reflect the proportion of voters who are African American, or your predictions will be entirely too rosy for Democratic candidates (since African Americans tend to vote overwhelmingly for Democrats, and whites do not). I discuss these issues in more detail in the next section.

Finally, note that the probability of selection into a sample should not depend on potential outcomes. Typically you stratify on the basis of explanatory variables. By doing so, and oversampling cases from the minority groups, you increase the variance on your explanatory variables (demonstrated in the next section). Because you can't explain a variable with a constant (chapter 2), it's a good idea to boost the variance on your explanatory variables. Increasing the variance of your explanatory variables is especially important when you have a small sample, as we now see.

Samples in Qualitative Studies

The third principle of sampling provides a nice bridge to the discussion of sampling in qualitative studies, since samples in qualitative studies generally are much smaller than samples in quantitative studies. All the sampling principles above apply to qualitative studies: You don't need to eat the whole ox; large populations do not require larger samples than smaller populations do; how cases are selected is more important than how many cases are selected; and you might need to stratify your sample.

What *is* different, usually, about qualitative studies is the limited number of cases studied. To understand what difference that might make regarding sampling strategy, we need first to identify more precisely the advantages of large samples over small samples. By identifying those advantages, we can think about sampling strategies that qualitative researchers might use to compensate for small sample size.

Unless the sample size is very small (say 50 cases or fewer), the chief limitation of a small sample is *not* that it fails to provide a modestly reliable description of key features of a population. A small representative sample of 100 generally yields fairly accurate *point estimates* (to use statistical parlance) of a population, where that term refers to a sample statistic that summarizes some important population characteristic, such as average income, the ratio of men to women, the percentage married, or (as in the George Wallace example) the percentage of the electorate who will vote for a particular candidate.

Instead, the chief limitation of a small sample is its lack of analytic power—its lack of leverage for separating the effects of multiple causes. The problem is especially acute when causes have little variance and (as is

most often the case) covary with other causes or correlates of the dependent variable. In general, *the more highly correlated the causes, the more cases we need to separate their effects*.

To illustrate, suppose you want to test a theory about only-children. Your theory argues that children reared in a home without siblings grow up in a very different environment from that experienced by children with siblings, and that the difference in childhood environment implies different adult outcomes—perhaps only-children tend to be more self-confident as adults, and more likely to assume leadership positions. You want to examine this theory by doing a qualitative study of 100 American adults, some of whom have siblings and some of whom do not. We assume that $N = 100$ is a fixed upper limit—time and budget constraints will not allow you to collect data on more than 100 individuals total.

One issue you must worry about is obtaining enough variance on your explanatory variable. The problem arises because only about 5 percent of American adults are only-children. So in a random sample of 100 adults, you expect about 5 only-children—hardly enough to make confident inferences about the differences between only-children and others. The solution, as before, is to use stratified sampling: Divide the population into those who are only-children and those who are not, and randomly select 50 adults from each group. In other words, you deliberately oversample only-children. In that way you increase the variance on your independent variable (let's call it **no-sibs**) more than five-fold, from 0.0475 to 0.25 (the variance for a dichotomy is $\pi(1-\pi)$, where π is proportion in one of the categories).

By increasing the number of only-children from 5 to 50, you increase the reliability of your estimates of Y (your dependent variable) for only-children. Of course, you also reduce the reliability of your estimates of Y for others, since you have reduced the sample of others from 95 down to 50. But the gain in reliability in moving from 5 to 50 cases more than offsets the loss in reliability in moving from 95 to 50. You can have more confidence in your estimates of *group differences* where the groups consist of 50 each than in estimates where $N = 5$ for one group and $N = 95$ for the other.

But, you ask, does not oversampling fly in the face of the notion of representative sampling? It was the oversampling of Republicans that led to the infamous *Literary Digest* blunder on the 1936 presidential election. How are we to reconcile the advice here to oversample with the earlier discussion about the importance of using representative samples?

The key is to note that the aim of the *Literary Digest* poll was to draw inferences about a single population (voters in the 1936 election), whereas the aim of the only-child research is to draw inferences about *differences* between two populations, a majority population (adults who had siblings)

and a minority population (adults who were only-children). For the best estimates of differences between populations, we want a representative sample of cases from each population. The sample does not need to be representative for the populations merged unless we want to draw inferences about that overarching population—then we need to reweight our samples to make them representative of the overarching population. To illustrate, imagine we find, in our 50/50 sample of only-children and others, that adults who were only-children attain 14 years of education on average while other adults average 12 years. Then our estimate of the educational difference is two years, but our estimate of the average education for all adults is $12(.95)+14(.05)=12.1$ years, reflecting the relative weights of the two groups in the overarching adult population.

In short, the use of stratified random sampling to oversample smaller populations is often strategic for comparing minority groups to majority groups, particularly in the case of qualitative research, where the total sample size may be modest. In the final analysis, though, purposive oversampling of this type cannot alleviate a fundamental problem with small samples: the lack of power to adjudicate competing explanations.[7] Consider alternative explanations for an association between the variable **no-sibs** (no siblings) and some outcome variable Y. A birth-order theorist, for example, might argue that any only-child effect that you find is a first-born effect in disguise. In other words, what matters is not the absence of siblings, but being born first. And **no-sibs** is inherently correlated with the variable **first-born**, since only-children are also first-born children. The correlation nonetheless is not perfect, since some first-born children are only-children and some are not—so it *is* possible to separate the birth-order effect from the only-child effect empirically, at least in principle. But to do that we need to divide our sample further into three groups, not two—only-children, first born among those with siblings, not first born—and compare the outcome variable Y for the three groups. If it is **no-sibs** that matters for the outcome, then Y for the only-children group should differ from Y for the other two groups. On the other hand, if it is **first-born** that matters for the outcome variable, then the difference should be between the first born and others: Y should be the same for first born with or without siblings, and differ for those who are not first born.

---

[7] Some methodologists and philosophers of social science conceptualize cause as deterministic rather than probabilistic, so a purported cause can be eliminated by a single exception. If so, then small samples could be analytically powerful, since each case has the potential to eliminate a different explanation for some phenomenon. In that tradition, as Lieberson (1991) points out, small N's can lead to "big conclusions." I do not pursue that approach here because I believe there is an alternative view of causality that holds more promise for social research (see rule 5).

The important point here is the demand placed on data by the need to examine alternative explanations, even when we stratify on explanatory variables. The more explanations there are, the more such comparisons we need to make; the more comparisons we make, the more finely we must stratify; the more finely we stratify, the smaller the subsamples upon which our comparisons are made. The solution to this dilemma is a large overall sample. That is what I mean when I say that a larger sample has greater "analytic power"—a larger sample permits more comparisons, and with more comparisons you can investigate more explanations.

Without the luxury of a large sample to provide the basis for one comparison after another, qualitative researchers very often must decide which comparisons are strategic, and select a sample accordingly. It helps to stratify on explanatory variables and oversample the smaller populations within those strata, but a stratified sample helps only up to a point: In the final analysis there are limits to the number of strata you can use when the overall sample is small. As a result, qualitative studies generally are not well-suited for omnibus tests of many different explanations for some observed association. Qualitative methods *are* well-suited for providing thick description that can help place quantitative results in proper context. Qualitative methods sometimes can also provide precise critical tests—tests that often are difficult with conventional quantitative approaches. Indeed, through strategic data selection, qualitative approaches often extend or correct quantitative studies, as we discover subsequently with examples of rule 3, "Build reality checks into your research."

Is Meaningful Social Research Possible?

October 7, 1903, is famous in aviation history for the test that failed. As news reporters watched expectantly, test pilot Charles Manly assumed his position aboard a curious-looking airship perched on a houseboat in the Potomac River. To avoid fatal landings, the plan was to crash into the water after demonstrating the possibility of human flight in a heavier-than-air machine. As propeller wheels whirred a foot from Manly's head, a mechanic cut the cable holding the catapult, and the airship tumbled ignominiously into the waters of the Potomac.

That spectacular failure, along with a second well-publicized plunge into the icy Potomac two months later, resulted in public ridicule of the airship's designer, the distinguished physicist and astronomer Samuel Pierpont Langley. Armed with a $50,000 grant from the War Department and another $20,000 from the Smithsonian Institution (which combined would be worth over $1.5 million in today's money), Langley had set out to build a heavier-than-air aircraft piloted by humans. His highly visible

failures added credibility to Lord Kelvin's famous claim, eight years earlier, that "Heavier-than-air flying machines are impossible."

After Langley's first failed attempt a *New York Times* editorial had this to say about the prospects of human flight ("Flying Machines Which Do Not Fly," *New York Times*, October 9, 1903):

> [It] might be assumed that the flying machine which will really fly might be evolved by the combined and continuous efforts of mathematicians and mechanicians in from one million to ten million years. . . . No doubt the problem has attractions for those it interests, but to the ordinary man it would seem as if effort might be employed more profitably.

In one of the ironies of history, as these words were being penned Wilbur and Orville Wright were preparing their own launch in a remote area on the coast of North Carolina. The first manned flight did not take place a million years later, but just a little more than two months after the *Times* editorial. On December 17, 1903, Orville and Wilbur took turns piloting a flying machine (that cost about $1,000 to build—$22,000 in today's money) over the sand dunes of Kitty Hawk. With no government financial backing and little fanfare, two determined bicycle builders from Ohio had proved Lord Kelvin and the *New York Times* wrong.

In academia there are Lord Kelvins among us today who make similar assertions about the impossibility of meaningful social research. Let me be very clear: I am not talking about those who say that good social research is difficult, or that social research is often badly done, or that some questions about human behavior are inherently unanswerable. (I happen to agree with all those assertions.) Rather, I am talking about those who are skeptical about the possibility of discovering general principles of human behavior. In the view of such skeptics, either such principles don't exist, or the principles are so "contextualized" that we cannot *generalize* from situation to situation. Empirical nihilism of this sort tends to be most prominent in schools of thought containing a "post" prefix: postmodernism, postpositivism, poststructuralism, post-Kuhnian philosophy of science, and so on.

In its most extreme version, empirical nihilism in the social sciences denies the possibility of discovering even regularities in human behavior. That position is obviously silly. Consider the life insurance industry. Actuarial tables presume knowledge about regularities in death rates. Although we cannot predict when specific individuals will die, we do know that women tend to outlive men (in most Western countries anyway), that nonsmokers tend to outlive smokers, and that it's good to exercise and eat your vegetables. Or consider government efforts to monitor economic growth and control inflation. If the discovery of general principles bearing

on economic behavior were beyond the ken of humans, the Federal Reserve Board would have no reason to meet to discuss alterations in the prime rate.

Empirical nihilism is not new, nor is it unique to the social sciences. Before attempting to discover principles about the world of nature, physicists and other natural scientists presuppose that observable natural phenomena, such as motion, are real, and that they are amenable to investigation. The ancient Greek philosopher Zeno of Elea believed otherwise. Zeno, like his teacher Parmenides, viewed the world of sense as an illusion. To defend that view, Zeno used paradoxes to demonstrate the impossibility of motion. One paradox can by illustrated by imagining an arrow in flight. At each moment in time, the arrow is located at a specific position. In subsequent moments, the arrow is at rest for the same reason. So the arrow is at rest at all moments. Thus if time is a continuous series of moments there is no time for the arrow to move, and motion is impossible. The paradoxes are challenging—they have engaged philosophers and others since they were introduced by Zeno almost 2,500 years ago—and important, since they undermine the conventional view that the sensory world is real. (The arrow paradox and other paradoxes of motion credited to Zeno are generally thought to have been resolved by calculus and Newtonian physics, although quantum mechanics has complicated that resolution in some instances.)

Is the social world real? Social science would be a strange choice of vocation for someone who answered that question negatively. Yet if we press the social contextualization argument far enough we arrive at an apparent denial of the social world. Substitute the term "context" for the term "moment in time" and you see that the contextualization argument is similar to Zeno's argument about the arrow. In effect Zeno viewed each moment as a separate context and concluded that the physical world is an illusion. If social acts are so radically contextualized that there are no general principles to discover, then one wonders what is the glue that holds human societies together. The existence of routine, predictable human interaction suggests that there is something "there" in the social world, even as the existence of motion suggests that there is something "there" in the physical world.

The best response to empirical nihilism is to ignore it and do the research. The aim of this book is to help both novice and experienced researchers "do the research." I have tried to make the book eminently practical. I provide examples of good social research, as well as examples of the other kind. Obviously I believe that meaningful social research *is* possible—imperfect research, but research that can be valuable nonetheless. (I say "can be valuable" because the findings of social research are subject to misuse. In the worst-case scenario, social science could be used

by authoritarian governments for the purpose of large-scale social engineering, which generally has been a miserable failure.) The pitfalls of social research may be many, but so too are the potential benefits. The seven rules are intended to move us closer to that potential in our social research projects.

Summary

There should be the possibility of surprise in social research. Social research requires that researchers remain open to new evidence, whether or not that evidence supports their cherished theories or beliefs. This openness leads to the possibility of surprise. Openness also makes research more exciting. If we know the results beforehand, what is the impetus for the research?

An important first step for effective research is choosing an appropriate research question—one where surprise is possible. Research questions should be both researchable and interesting. Research questions are uninteresting when there is no possibility of learning anything new, when there is no perceived relevance to theory or everyday life, or when the answer is significant only to a very small number of people.

Because important social issues of the day typically have been investigated in one form or another by social scientists, most often research evidence already exists on the question you want to investigate. In that case you need to design your research to ensure that surprise is possible. The first step is to distinguish well-established findings from the other kind. By "the other kind" I mean findings that are mixed or inconclusive, or findings that are new and have not yet been replicated in subsequent studies. Surprise *is* possible in those instances, because in those instances there are lingering questions about earlier results. Hence your research, if properly executed, contributes to knowledge.

By contrast, if prior findings are widely viewed as settled, you contribute little or nothing by merely repeating the analyses of others. Instead you must either challenge the earlier findings or extend/clarify them. If there is a good reason to doubt the results of prior studies, then of course it is worthwhile to challenge them with new research. In that case, though, the burden of proof rests on you to identify the flaws in prior research that are serious enough to cast doubt on their findings (flaws that you will correct in your research). The other strategy—to extend or clarify earlier findings—can be implemented by examining results for a subpopulation or for a different population or time period. New understanding of old findings can also be gained by introducing mediating variables or moderating variables or nonmediating variables (or some combination of

all three), or by testing for reverse causation. In some instances a synthesizing study may be possible by bridging prior research in different subfields.

After deciding on a research question, you must decide whom to study. First you decide what population you want to examine. Then you select a sample of that population since generally you will lack resources to collect data on the entire population. The important feature of a sample is not its size (within reason) but its representativeness. When samples are properly selected, it is possible to draw defensible conclusions about large populations (for example, all U.S. citizens) from analyses of only a tiny fraction of the population. These underlying sampling principles are the same for qualitative and quantitative research. Qualitative research, however, usually relies on smaller (sometimes much smaller) samples. Stratified random sampling that purposely oversamples smaller populations might compensate for the limited sample size, particularly if the aim is to test an uncomplicated model. A small sample does, however, limit your ability to make multiple comparisons with different variables. For this reason, qualitative studies generally are ill-suited for providing omnibus tests of multiple explanations for some observed association.