Posing problems properly is often more difficult than answering them. Indeed, a properly phrased question often seems to answer itself. You may have discovered the answer to a question just in the process of making the question clear to someone else.

Part 2 deals with what should be observed; that is, Part 2 considers the posing of proper scientific questions, the structuring of inquiry. Part 3 will describe some of the specific methods of social scientific observation.

Chapter 4 addresses the beginnings of research. It examines some of the purposes of inquiry, units of analysis, and the reasons scientists get involved in research projects.

Chapter 5 deals with the specification of what it is you want to measure—the processes of conceptualization and operationalization. It looks at some of the terms that you and I use quite casually in everyday life—*prejudice, liberalism, happiness,* and so forth—and

The Structuring of Inquiry

shows how essential it is to clarify what we really mean by such terms when we do research. This process of clarification is called conceptualization.

Once we clarify what we mean by certain terms, we can then measure the referents of those terms. The process of devising steps or operations for measuring what we want to study is called operationalization. Chapter 5 deals with the topic of operationalization in general, paying special attention to the framing of questions for interviews and questionnaires.

To complete the introduction to measurement, Chapter 6 breaks with the chronological discussion of how research is conducted. In this chapter, we'll examine techniques for measuring variables in quantitative research through the combination of several indicators: indexes, scales, and typologies. As an example, we might ask survey respondents five different questions about their attitudes toward gender equality and then combine the answers to all five questions into a composite measure of gender-based egalitarianism. Although such composite measures are constructed during the analysis of data (see Part 4), the raw materials for them must be provided for in the design and execution of data collection.

Finally, we'll look at how social researchers select people or things for observation. Chapter 7, on sampling, addresses the fundamental scientific issue of generalizability. As you'll see, we can select a few people or things for observation and then apply what we observe to a much larger group. For example, by surveying 2,000 U.S. citizens about whom they favor for president of the United States, we can accurately predict how tens of millions will vote. In this chapter, we'll examine techniques that increase the generalizability of what we observe.

What you learn in Part 2 will bring you to the verge of making controlled social scientific observations. Part 3 will then show you how to take that next step.

CHAPTER FOUR

Research Design

CHAPTER OVERVIEW

Here you'll see the wide variety of research designs available to social researchers as well as how to design a study—that is, specifying exactly who or what is to be studied when, how, and for what purpose.

SociologyNow™: Research Methods

Use this online tool to help you make the grade on your next exam. After reading this chapter, go to the "Online Study Resources" at the end of the chapter for instructions on how to benefit from *SociologyNow: Research Methods*.

Introduction

Science is an enterprise dedicated to "finding out." No matter what you want to find out, though, there will likely be a great many ways of doing it. That's true in life generally. Suppose, for example, that you want to find out whether a particular automobile—say, the new Burpo-Blasto—would be a good car for you. You could, of course, buy one and find out that way. Or you could talk to a lot of B-B owners or to people who considered buying one but didn't. You might check the classified ads to see if there are a lot of B-Bs being sold cheap. You could read a consumer magazine evaluation of Burpo-Blastos. A similar situation occurs in scientific inquiry.

Ultimately, scientific inquiry comes down to making observations and interpreting what you've observed, the subjects of Parts 3 and 4 of this book. Before you can observe and analyze, however, you need a plan. You need to determine what you're going to observe and analyze: why and how. That's what research design is all about.

Although the details vary according to what you wish to study, you face two major tasks in any research design. First, you must specify as clearly as possible what you want to find out. Second, you must determine the best way to do it. Interestingly, if you can handle the first consideration fully, you'll probably handle the second in the same process. As mathematicians say, a properly framed question contains the answer.

Let's say you're interested in conducting social research on terrorism. When Jeffrey Ross (2004) addressed this issue, he found the existing studies used a variety of qualitative and quantitative approaches. Qualitative researchers, for example, generated original data through

> Autobiographies
>
> Incident Reports and Accounts
>
> Hostages' Experiences with Terrorists
>
> Firsthand Accounts of Implementing Policies

Ross goes on to discuss some of the secondary materials used by qualitative researchers: "biographies of terrorists, case studies of terrorist organizations, case studies on types of terrorism, case studies on particular terrorist incidents, and case studies of terrorism in selected regions and countries." (2004: 27) Quantitative researchers, on the other hand, addressed terrorism in a variety of ways, including analyses of media coverage, statistical modeling of terrorist events, and the use of various databases relevant to the topic. As you'll see in this chapter, any research topic can be approached from many different directions.

This chapter provides a general introduction to research design, whereas the other chapters in Part 2 elaborate on specific aspects of it. In practice, all aspects of research design are interrelated. As you read through Part 2, the interrelationships among parts will become clearer.

We'll start by briefly examining the main purposes of social research. Then, we'll consider units of analysis—the what or whom you want to study. Next we'll consider ways of handling time in social research, or how to study a moving target that changes over time.

With these ideas in hand, we'll turn to how to design a research project. This overview of the research process serves two purposes: Besides describing how you might go about designing a study, it provides a map of the remainder of this book.

Finally, we'll look at the elements of research proposals. Often, the actual conduct of research needs to be preceded by a detailing of your intentions—to obtain funding for a major project or perhaps to get your instructor's approval for a class project. You'll see that the research proposal provides an excellent opportunity for you to consider all aspects of your research in advance.

Three Purposes of Research

Social research can serve many purposes. Three of the most common and useful purposes are exploration, description, and explanation. Although a given study can have more than one of these purposes—and most do—examining them

separately is useful because each has different implications for other aspects of research design.

Exploration

Much of social research is conducted to explore a topic, that is, to start to familiarize a researcher with that topic. This approach typically occurs when a researcher examines a new interest or when the subject of study itself is relatively new.

As an example, let's suppose that widespread taxpayer dissatisfaction with the government erupts into a taxpayers' revolt. People begin refusing to pay their taxes, and they organize themselves around that issue. You might like to learn more about the movement: How widespread is it? What levels and degrees of support are there within the community? How is the movement organized? What kinds of people are active in it? An exploratory study could help you find at least approximate answers to some of these questions. You might check figures with tax-collecting officials, collect and study the literature of the movement, attend meetings, and interview leaders.

Exploratory studies are also appropriate for more persistent phenomena. Suppose you're unhappy with your college's graduation requirements and want to help change them. You might study the history of such requirements at the college and meet with college officials to learn the reasons for the current standards. You could talk to several students to get a rough idea of their sentiments on the subject. Though this last activity would not necessarily yield an accurate picture of student opinion, it could suggest what the results of a more extensive study might be.

Sometimes exploratory research is pursued through the use of focus groups, or guided small-group discussions. This technique is frequently used in market research; we'll examine it further in Chapter 10.

Exploratory studies are most typically done for three purposes: (1) to satisfy the researcher's curiosity and desire for better understanding, (2) to test the feasibility of undertaking a more extensive study, and (3) to develop the methods to be employed in any subsequent study.

A while back, for example, I became aware of the growing popularity of something called "channeling," in which a person known as a channel or medium enters a trance state and begins speaking with a voice that claims it originates outside the channel. Some of the voices say they come from a spirit world of the dead, some say they are from other planets, and still others say they exist in dimensions of reality difficult to explain in ordinary human terms.

The channeled voices, often referred to as entities, sometimes use the metaphor of radio or television for the phenomenon they represent. "When you watch the news," one told me in the course of an interview, "you don't believe Dan Rather is really inside the television set. The same is true of me. I use this medium's body the way Dan Rather uses your television set."

The idea of channeling interested me from several perspectives, not the least of which was the methodological question of how to study scientifically something that violates so much of what we take for granted, including scientific staples such as space, time, causation, and individuality.

Lacking any rigorous theory or precise expectations, I merely set out to learn more. Using some of the techniques of qualitative field research discussed in Chapter 10, I began amassing information and forming categories for making sense of what I observed. I read books and articles about the phenomenon and talked to people who had attended channeling sessions. I then attended channeling sessions myself, observing those who attended as well as the channel and entity. Next, I conducted personal interviews with numerous channels and entities.

In most interviews, I began by asking the human channels questions about how they first began channeling, what it was like, and why they continued, as well as standard biographical questions. The channel would then go into a trance, whereby the interview continued with the entity speaking. "Who are you?" I might ask. "Where do you come from?" "Why are you doing this?" "How can I tell if you are real or a fake?" Although I went into these interview sessions with several questions prepared in advance, each of the interviews followed

whatever course seemed appropriate in light of the answers given.

This example of exploration illustrates where social research often begins. Whereas researchers working from deductive theories have the key variables laid out in advance, one of my first tasks was to identify some of the possibly relevant variables. For example, I noted a channel's gender, age, education, religious background, regional origins, and previous participation in things metaphysical. I chose most of these variables because they commonly affect behavior.

I also noted differences in the circumstances of channeling sessions. Some channels said they must go into deep trances, some use light trances, and others remain conscious. Most sit down while channeling, but others stand and walk about. Some channels operate under pretty ordinary conditions; others seem to require metaphysical props such as dim lights, incense, and chanting. Many of these differences became apparent to me only in the course of my initial observations.

Regarding the entities, I have been interested in classifying where they say they come from. Over the course of my interviews, I've developed a set of questions about specific aspects of "reality," attempting to classify the answers they give. Similarly, I ask each to speak about future events.

Over the course of this research, my examination of specific topics has become increasingly focused as I've identified variables that seem worth pursuing: gender, education, and religion, for example. Note, however, that I began with a reasonably blank slate.

Exploratory studies are quite valuable in social scientific research. They're essential whenever a researcher is breaking new ground, and they almost always yield new insights into a topic for research. Exploratory studies are also a source of grounded theory, as discussed in Chapter 2.

The chief shortcoming of exploratory studies is that they seldom provide satisfactory answers to research questions, though they can hint at the answers and can suggest which research methods could provide definitive ones. The reason exploratory studies are seldom definitive in themselves has to do with representativeness; that is, the people you study in your exploratory research may not be typical of the larger population that interests you. Once you understand representativeness, you'll be able to know whether a given exploratory study actually answered its research problem or only pointed the way toward an answer. (Representativeness is discussed at length in Chapter 7.)

Description

A major purpose of many social scientific studies is to describe situations and events. The researcher observes and then describes what was observed. Because scientific observation is careful and deliberate, however, scientific descriptions are typically more accurate and precise than casual ones are.

The U.S. Census is an excellent example of descriptive social research. The goal of the census is to describe accurately and precisely a wide variety of characteristics of the U.S. population, as well as the populations of smaller areas such as states and counties. Other examples of descriptive studies are the computation of age-gender profiles of populations done by demographers, the computation of crime rates for different cities, and a product-marketing survey that describes the people who use, or would use, a particular product. A researcher who carefully chronicles the events that take place on a labor union picket line has, or at least serves, a descriptive purpose. A researcher who computes and reports the number of times individual legislators voted for or against organized labor also fulfills a descriptive purpose.

Many qualitative studies aim primarily at description. An anthropological ethnography, for example, may try to detail the particular culture of some preliterate society. At the same time, such studies are seldom limited to a merely descriptive purpose. Researchers usually go on to examine *why* the observed patterns exist and what they imply.

Explanation

The third general purpose of social scientific research is to explain things. Descriptive studies answer questions of what, where, when, and

how; explanatory questions, of why. So when William Sanders (1994) set about describing the varieties of gang violence, he also wanted to reconstruct the process that brought about violent episodes among the gangs of different ethnic groups.

Reporting the voting intentions of an electorate is descriptive, but reporting why some people plan to vote for Candidate A and others for Candidate B is explanatory. Identifying variables that explain why some cities have higher crime rates than others involves explanation. A researcher who sets out to know why an antiabortion demonstration ended in a violent confrontation with police, as opposed to simply describing what happened, has an explanatory purpose.

Let's look at a specific case. What factors do you suppose might shape people's attitudes toward the legalization of marijuana? To answer this, you might first consider whether men and women differ in their opinions. An explanatory analysis of the 2002 General Social Survey (GSS) data indicates that 38 percent of men and 30 percent of women said marijuana should be legalized.

What about political orientation? The GSS data show that 55 percent of liberals said marijuana should be legalized, compared with 29 percent of moderates and 27 percent of conservatives. Further, 41 percent of Democrats, compared with 34 percent of Independents and 28 percent of Republicans, supported legalization.

Given these statistics, you might begin to develop an explanation for attitudes toward marijuana legalization. Further study of gender and political orientation might then lead to a deeper explanation of these attitudes.

correlation An empirical relationship between two variables such that (1) changes in one are associated with changes in the other or (2) particular attributes of one variable are associated with particular attributes of the other. Correlation in and of itself does not constitute a causal relationship between the two variables, but it is one criterion of causality.

The Logic of Nomothetic Explanation

The preceding examination of what factors might cause attitudes about legalizing marijuana illustrates nomothetic explanation, as discussed in Chapter 1. Recall that in this model, we try to find a few factors (independent variables) that can account for many of the variations in a given phenomenon. This explanatory model stands in contrast to the idiographic model, in which we seek a complete, in-depth understanding of a single case.

In our example, an idiographic approach would suggest all the reasons that one person was opposed to legalization—involving what her parents, teachers, and clergy told her about it; any bad experiences experimenting with it; and so forth. When we understand something idiographically, we feel we *really* understand it. When we know all the reasons why someone opposed legalizing marijuana, we couldn't imagine that person having any other attitude.

In contrast, a nomothetic approach might suggest that overall political orientations account for much of the difference of opinion about legalizing marijuana. Because this model is inherently probabilistic, it is more open than the idiographic model to misunderstanding and misinterpretation. Let's examine what social researchers mean when they say one variable (nomothetically) causes another. Then, we'll look at what they *don't* mean.

Criteria for Nomothetic Causality

There are three main criteria for nomothetic causal relationships in social research: (1) the variables must be correlated, (2) the cause takes place before the effect, and (3) the variables are nonspurious.

Correlation

Unless some actual relationship—or **correlation**—is found between two variables, we can't say that a causal relationship exists. Our analysis of GSS data suggested that political orientation was a cause of attitudes about legalizing marijuana. Had the same percentage of liberals and conservatives supported

legalization, we could hardly say that political orientations caused the attitude. Though this criterion is obvious, it emphasizes the need to base social research assertions on actual observations rather than assumptions.

Time Order

Next, we can't say a causal relationship exists unless the cause precedes the effect in time. Notice that it makes more sense to say that most children's religious affiliations are caused by those of their parents than to say that parents' affiliations are caused by those of their children—even though it would be possible for you to change your religion and for your parents to follow suit. Remember, nomothetic explanation deals with "most cases" but not all.

In our marijuana example, it would make sense to say that gender causes, to some extent, attitudes toward legalization, whereas it would make no sense to say that opinions about marijuana determine a person's gender. Notice, however, that the time order connecting political orientations and attitudes about legalization is less clear, though we sometimes reason that general orientations cause specific opinions. And sometimes our analyses involve two or more independent variables that were established at the same time: looking at the effects of gender and race on voting behavior, for example. As we'll see in the next chapter, the issue of time order can be a complex matter.

Nonspurious

The third requirement for a causal relationship is that the effect cannot be explained in terms of some third variable. For example, there is a correlation between ice-cream sales and deaths due to drowning: the more ice cream sold, the more drownings, and vice versa. There is, however, no direct link between ice cream and drowning. The third variable at work here is *season* or *temperature*. Most drowning deaths occur during summer—the peak period for ice-cream sales.

Here are a couple of other examples of **spurious relationships**, or ones that aren't genuine. There is a negative relationship between the number of mules and the number of Ph.D.'s in towns and cities: the more mules, the fewer Ph.D.'s and vice versa. Perhaps you can think of another variable that would explain this apparent relationship. The answer is rural versus urban settings. There are more mules (and fewer Ph.D.'s) in rural areas, whereas the opposite is true in cities.

Or, consider the positive correlation between shoe size and math ability among schoolchildren. Here, the third variable that explains the puzzling relationship is age. Older children have bigger feet and more highly developed math skills, on average, than younger children do. See Figure 4-1 for an illustration of this spurious relationship. Observed associations are indicated with thin arrows; causal relationships with thick ones. Notice that observed associations go in both directions. That is, as one variable occurs or changes, so does the other.

The list goes on. Areas with many storks have high birth rates. Those with few storks have low birth rates. Do storks really deliver babies? Birth rates are higher in the country than in the city; more storks live in the country than the city. The third variable here is *urban/rural areas*.

Finally, the more fire trucks that put out a fire, the more damage to the structure. Can you guess what the third variable is? In this case, it's the *size of the fire*.

Thus, when social researchers say there is a causal relationship between, say, education and racial tolerance, they mean (1) there is a statistical correlation between the two variables, (2) a person's educational level occurred before their current level of tolerance or prejudice, and (3) there is no third variable that can explain away the observed correlation as spurious.

False Criteria for Nomothetic Causality

Because notions of cause and effect are well entrenched in everyday language and logic, it's important to specify some of the things social

spurious relationship A coincidental statistical correlation between two variables, shown to be caused by some third variable.

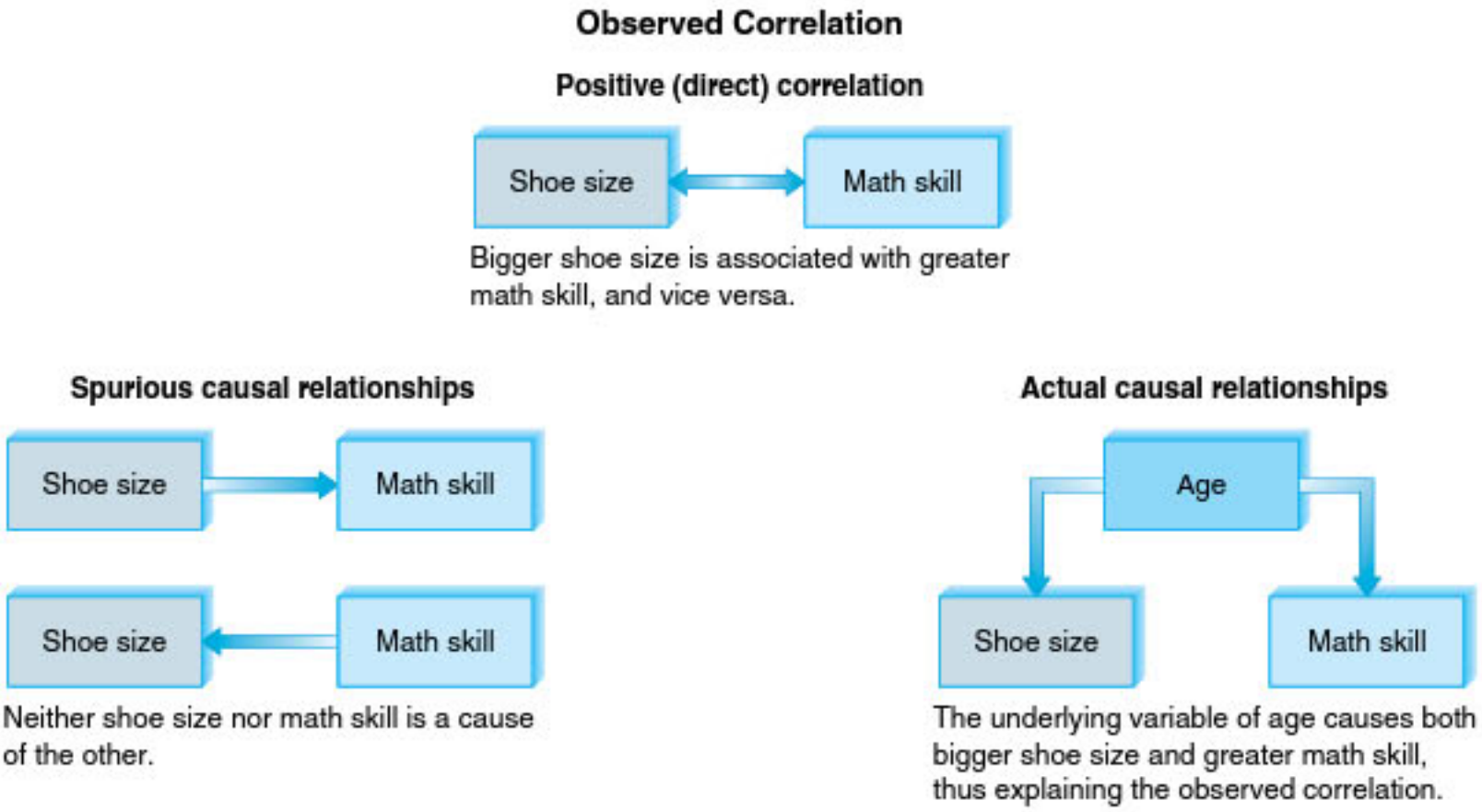


FIGURE 4-1

An Example of a Spurious Causal Relationship. Finding an empirical correlation between two variables does not necessarily establish a causal relationship. Sometimes the observed correlation is the incidental result of other causal relationships, involving other variables.

researchers do *not* mean when they speak of causal relationships. When they say that one variable causes another, they do not necessarily mean to suggest complete causation, to account for exceptional cases, or to claim that the causation exists in a majority of cases.

Complete Causation

Whereas an idiographic explanation of causation is relatively complete, a nomothetic explanation is probabilistic and usually incomplete. As we've seen, social researchers may say that political orientations cause attitudes toward legalizing marijuana even though not all liberals approve nor all conservatives disapprove. Thus, we say that political orientation is one of the causes of the attitude, but not the only one.

Exceptional Cases

In nomothetic explanations, exceptions do not disprove a causal relationship. For example, it is consistently found that women are more religious than men in the United States. Thus, gender may be a cause of religiosity, even if your uncle is a religious zealot or you know a woman who is an avowed atheist. Those exceptional cases do not disprove the overall, causal pattern.

Majority of Cases

Causal relationships can be true even if they don't apply in a majority of cases. For example, we say that children who are not supervised after school are more likely to become delinquent than those who are supervised are; hence, lack of supervision is a cause of delinquency. This causal relationship holds true even if only a small percentage of those not supervised become delinquent. As long as they are *more likely* than those who are supervised to be delinquent, we say there is a causal relationship.

The social scientific view of causation may vary from what you are accustomed to, because people commonly use the term *cause* to mean something

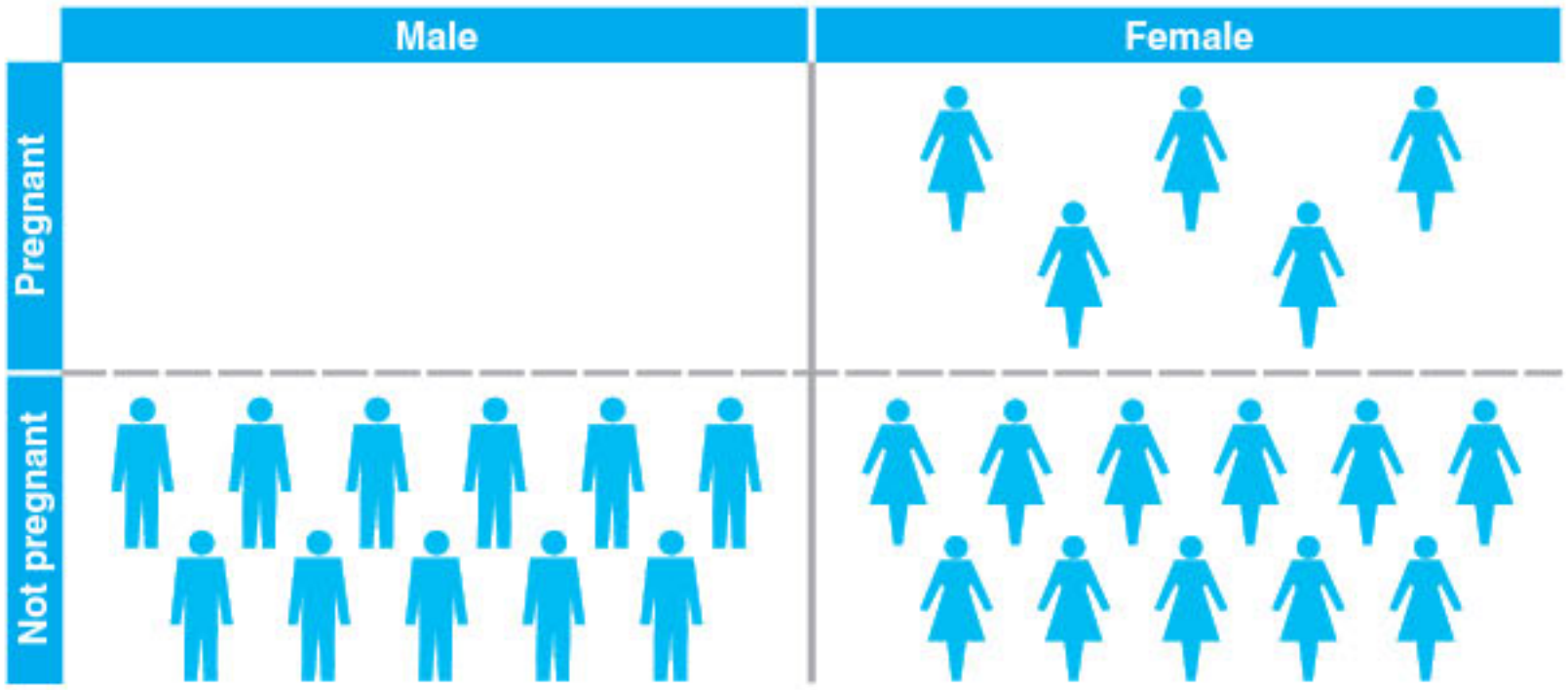


FIGURE 4-2

Necessary Cause. Being female is a necessary cause of pregnancy; that is, you can't get pregnant unless you are female.

that completely causes another thing. The somewhat different standard used by social researchers can be seen more clearly in terms of necessary and sufficient causes.

Necessary and Sufficient Causes

A *necessary cause* represents a condition that *must* be present for the effect to follow. For example, it is necessary for you to take college courses in order to get a degree. Take away the courses, and the degree never follows. However, simply taking the courses is not a sufficient cause of getting a degree. You need to take the right ones and pass them. Similarly, being female is a necessary condition of becoming pregnant, but it is not a sufficient cause. Otherwise, all women would get pregnant.

Figure 4-2 illustrates this relationship between the variables of gender and pregnancy as a matrix showing the possible outcomes of combining these variables.

A *sufficient cause*, on the other hand, represents a condition that, if it is present, guarantees the effect in question. This is not to say that a sufficient cause is the *only* possible cause of a particular effect. For example, skipping an exam in this course would be a sufficient cause for failing it, though students could fail it other ways as well. Thus, a cause can be sufficient, but not necessary. Figure 4-3 illustrates the relationship between taking or not taking the exam and either passing or failing it.

The discovery of a cause that is both necessary *and* sufficient is, of course, the most satisfying outcome in research. If juvenile delinquency were the effect under examination, it would be nice to discover a single condition that (1) must be present for delinquency to develop and (2) always results in delinquency. In such a case, you would surely feel that you knew precisely what caused juvenile delinquency.

Unfortunately, we never discover single causes that are absolutely necessary and absolutely sufficient when analyzing the nomothetic relationships among variables. It is not uncommon, however, to find causal factors that are either 100 percent necessary (you must be female to become pregnant) or 100 percent sufficient (skipping an exam will inevitably cause you to fail it).

In the idiographic analysis of single cases, you may reach a depth of explanation from which it is

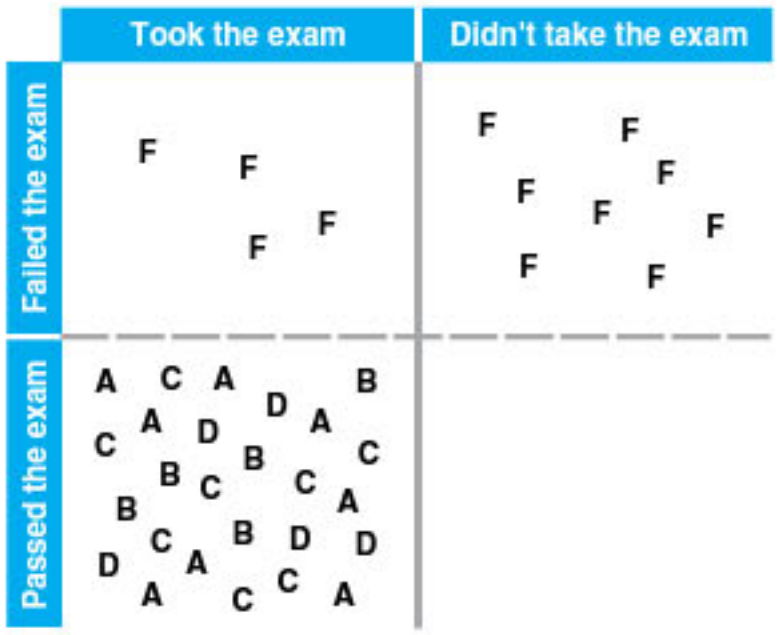


FIGURE 4-3

Sufficient Cause. Not taking the exam is a sufficient cause of failing it, even though there are other ways of failing (such as answering randomly).

reasonable to assume that things could not have turned out differently, suggesting you have determined the *sufficient* causes for a particular result. (Anyone with all the same details of your genetic inheritance, upbringing, and subsequent experiences would have ended up going to college.) At the same time, there could always be other causal paths to the same result. Thus, the idiographic causes are sufficient but not necessary.

Units of Analysis

In social research, there is virtually no limit to what or whom can be studied, or the **units of analysis**. This topic is relevant to all forms of social research, although its implications are clearest in the case of nomothetic, quantitative studies.

The idea for units of analysis may seem slippery at first, because research—especially nomothetic research—often studies large collections of people or things, or aggregates. It's important to distinguish between the unit of analysis and the aggregates that we generalize about. For instance, a researcher may study a class of people, such as Democrats, college undergraduates, African American women under 30, or some other collection. But if the researcher is interested in exploring, describing, or explaining how different groups of individuals behave *as individuals*, the unit of analysis is the individual, not the group. This is true even though the researcher uses the information about individuals to generalize about aggregates of individuals, as in saying that more Democrats than Republicans favor legalizing marijuana. Think of it this way: Having an attitude about marijuana is something that can only be an attribute of an individual, not a group; that is, there is no one group "mind" that can have an attitude. So even when we generalize about Democrats, we're generalizing about an attribute they possess as individuals.

In contrast, we may sometimes want to study groups, considered as individual "actors" or entities that have attributes as groups. For instance, we might want to compare the characteristics of different types of street gangs. In that case our unit of analysis would be gangs (not members of gangs), and we might proceed to make generalizations about different types of gangs. For example, we might conclude that male gangs are more violent than female gangs. Each gang (unit of analysis) would be described in terms of two variables: (1) What sex are the members? and (2) How violent are its activities? So we might study 52 gangs, reporting that 40 were male and 12 were female, and so forth. The "gang" would be the unit of analysis, even though some of the characteristics were drawn from the components (members) of the gangs.

Social researchers tend to choose individual people as their units of analysis. You may note the characteristics of individual people—gender, age, region of birth, attitudes, and so forth. You can then combine these descriptions to provide a composite picture of the group the individuals represent, whether a street-corner gang or a whole society.

For example, you may note the age and gender of each student enrolled in Political Science 110 and then characterize the group of students as

units of analysis The what or whom being studied. In social science research, the most typical units of analysis are individual people.

being 53 percent men and 47 percent women and as having a mean age of 18.6 years. Although the final description would be of the class as a whole, the description is based on characteristics that members of the class have as individuals.

The same distinction between units of analysis and aggregates occurs in explanatory studies. Suppose you wished to discover whether students with good study habits received better grades in Political Science 110 than students with poor study habits did. You would operationalize the variable *study habits* and measure this variable, perhaps in terms of hours of study per week. You might then aggregate students with good study habits and those with poor study habits and see which group received the best grades in the course. The purpose of the study would be to explain why some groups of students do better in the course than others do, but the unit of analysis is still individual students.

Units of analysis in a study are usually also the units of observation. Thus, to study success in a political science course, we would observe individual students. Sometimes, however, we "observe" our units of analysis indirectly. For example, suppose we want to find out whether disagreements about the death penalty tend to cause divorce. In this case, we might "observe" individual husbands and wives by asking them about their attitudes about capital punishment, in order to distinguish couples who agree and disagree on this issue. In this case, our units of observation are individual wives and husbands, but our units of analysis (the things we want to study) are couples.

Units of analysis, then, are those things we examine in order to create summary descriptions of all such units and to explain differences among them. In most research projects, the unit of analysis will probably be clear to you. When the unit of analysis is not clear, however, it's essential to determine what it is; otherwise, you cannot determine what observations are to be made about whom or what.

Some studies try to describe or explain more than one unit of analysis. In these cases, the researcher must anticipate what conclusions she or he wishes to draw with regard to which units of analysis. For example, we may want to discover what kinds of college students (individuals) are most successful in their careers; we may also want to learn what kinds of colleges (organizations) produce the most successful graduates.

Here's an example that illustrates the complexity of units of analysis. Murder is a fairly personal matter: One individual kills another individual. However, when Charis Kubrin and Ronald Weitzer (2003: 157) ask, "Why do these neighborhoods generate high homicide rates?" the unit of analysis in that phrase is *neighborhood*. You can probably imagine some kinds of neighborhoods (e.g., poor, urban) that would have high homicide rates and some (e.g., wealthy, suburban) that would have low rates. In this particular conversation, the unit of analysis (neighborhood) would be categorized in terms of variables such as *economic level*, *locale*, and *homicide rate*.

In their analysis, however, Kubrin and Weitzer were also interested in different types of homicide: in particular, those that occurred in retaliation for some earlier event, such as an assault or insult. Can you identify the unit of analysis common to all of the following excerpts?

1. The sample of killings . . .
2. The coding instrument includes over 80 items related to the homicide.
3. Of the 2,161 homicides that occurred from 1985 [to] 1995 . . .
4. Of those with an identified motive, 19.5 percent (n = 337) are retaliatory.

(Kubrin and Weitzer 2003: 163)

In each of these excerpts, the unit of analysis is *homicide* (also called killing or murder). Sometimes you can identify the unit of analysis in the description of the sampling methods, as in the first excerpt. A discussion of classification methods might also identify the unit of analysis, as in the second excerpt (80 ways to code the homicides). Often, numerical summaries point the way: 2,161 homicides; 19.5 percent (of the homicides). With a little practice you'll be able to identify the units of analysis in most social research reports, even when more than one is used in a given analysis.

To explore this topic in more depth, let's consider several common units of analysis in social research.

Individuals

As mentioned, individual human beings are perhaps the most typical units of analysis for social research. Social researchers tend to describe and explain social groups and interactions by aggregating and manipulating the descriptions of individuals.

Any type of individual may be the unit of analysis for social research. This point is more important than it may seem at first. The norm of generalized understanding in social research should suggest that scientific findings are most valuable when they apply to all kinds of people. In practice, however, social researchers seldom study all kinds of people. At the very least, their studies are typically limited to the people living in a single country, though some comparative studies stretch across national boundaries. Often, though, studies are quite circumscribed.

Examples of classes of individuals that might be chosen for study include students, gays and lesbians, auto workers, voters, single parents, and faculty members. Note that each of these terms implies some population of individuals. Descriptive studies with individuals as their units of analysis typically aim to describe the population that comprises those individuals, whereas explanatory studies aim to discover the social dynamics operating within that population.

As the units of analysis, individuals may be characterized in terms of their membership in social groupings. Thus, an individual may be described as belonging to a rich family or to a poor one, or a person may be described as having a college-educated mother or not. We might examine in a research project whether people with college-educated mothers are more likely to attend college than are those with non-college-educated mothers or whether high school graduates in rich families are more likely than those in poor families to attend college. In each case, the unit of analysis—the "thing" whose characteristics we are seeking to describe or explain—is the individual. We then aggregate these individuals and make generalizations about the population they belong to.

Groups

Social groups can also be units of analysis in social research. That is, we may be interested in characteristics that belong to one group, considered as a single entity. If you were to study the members of a criminal gang to learn about criminals, the individual (criminal) would be the unit of analysis; but if you studied all the gangs in a city to learn the differences, say, between big gangs and small ones, between "uptown" and "downtown" gangs, and so forth, you would be interested in gangs rather than their individual members. In this case, the unit of analysis would be the gang, a social group.

Here's another example. Suppose you were interested in the question of access to computers in different segments of society. You might describe families in terms of total annual income and according to whether or not they had computers. You could then aggregate families and describe the mean income of families and the percentage with computers. You would then be in a position to determine whether families with higher incomes were more likely to have computers than were those with lower incomes. In this case, the unit of analysis would be families.

As with other units of analysis, we can derive the characteristics of social groups from those of their individual members. Thus, we might describe a family in terms of the age, race, or education of its head. In a descriptive study, we might find the percentage of all families that have a college-educated head of family. In an explanatory study, we might determine whether such families have, on average, more or fewer children than do families headed by people who have not graduated from college. In each of these examples, the family is the unit of analysis. In contrast, had we asked whether college-educated individuals have more or fewer children than do their less-educated counterparts, then the individual would have been the unit of analysis.

Other units of analysis at the group level could be friendship cliques, married couples, census

blocks, cities, or geographic regions. As with individuals, each of these terms implies some population. *Street gangs* implies some population that includes all street gangs, perhaps in a given city. You might then describe this population by generalizing from your findings about individual gangs. For instance, you might describe the geographic distribution of gangs throughout a city. In an explanatory study of street gangs, you might discover whether large gangs are more likely than small ones to engage in intergang warfare. Thus, you would arrive at conclusions about the population of gangs by using individual groups as your unit of analysis.

Organizations

Formal social organizations may also be the units of analysis in social research. For example, a researcher might study corporations, by which he or she implies a population of all corporations. Individual corporations might be characterized in terms of their number of employees, net annual profits, gross assets, number of defense contracts, percentage of employees from racial or ethnic minority groups, and so forth. We might determine whether large corporations hire a larger or smaller percentage of minority group employees than do small corporations. Other examples of formal social organizations suitable as units of analysis include church congregations, colleges, army divisions, academic departments, and supermarkets.

Figure 4-4 provides a graphic illustration of some different units of analysis and the statements that might be made about them.

Social Interactions

Sometimes social interactions are the relevant units of analysis. Instead of individual humans, you can study what goes on between them: telephone calls, kisses, dancing, arguments, fistfights, e-mail exchanges, chat-room discussions, and so forth. As you saw in Chapter 2, social interaction is the basis for one of the primary theoretical paradigms in the social sciences, and the number of units of analysis that social interactions provide is nearly infinite.

Even though individuals are usually the actors in social interactions, there is a difference between (1) comparing the kinds of people who subscribe to different Internet service providers (individuals being the unit of analysis) and (2) comparing the length of chat-room discussions on those same ISPs (the discussion being the unit of analysis).

Social Artifacts

Another unit of analysis is the **social artifact**, or any product of social beings or their behavior. One class of artifacts includes concrete objects such as books, poems, paintings, automobiles, buildings, songs, pottery, jokes, student excuses for missing exams, and scientific discoveries.

For example, Lenore Weitzman and her associates (1972) were interested in learning how gender roles are taught. They chose children's picture books as their unit of analysis. Specifically, they examined books that had received the Caldecott Medal. Their results were as follows:

> We found that females were underrepresented in the titles, central roles, pictures, and stories of every sample of books we examined. Most children's books are about boys, men, male animals, and deal exclusively with male adventures. Most pictures show men singly or in groups. Even when women can be found in the books, they often play insignificant roles, remaining both inconspicuous and nameless.
>
> *(Weitzman et al. 1972: 1128)*

In a more recent study, Roger Clark, Rachel Lennon, and Leana Morris (1993) concluded that male and female characters are now portrayed less stereotypically than before, observing a clear progress toward portraying men and women in nontraditional roles. However, they did not find total equality between the sexes.

As this example suggests, just as people or social groups imply populations, each social object

social artifact Any product of social beings or their behavior. Can be a unit of analysis.

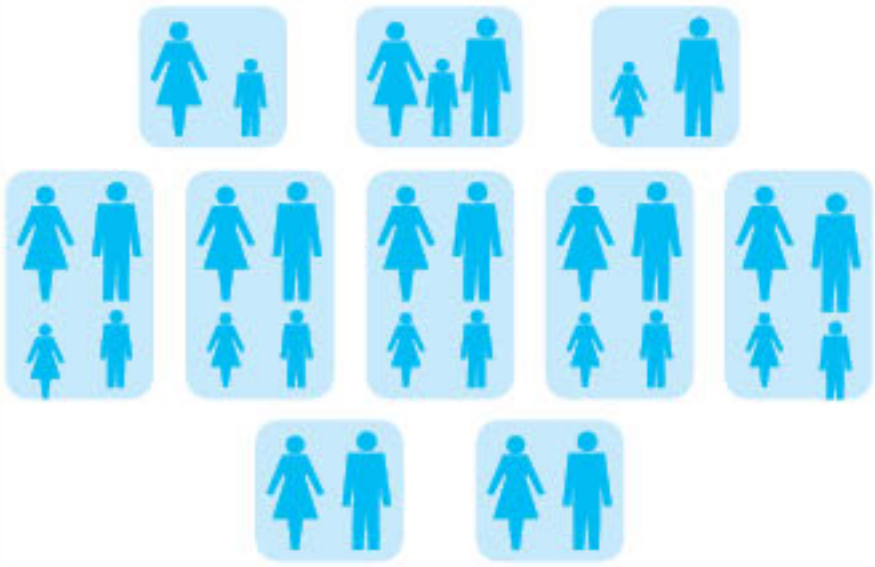
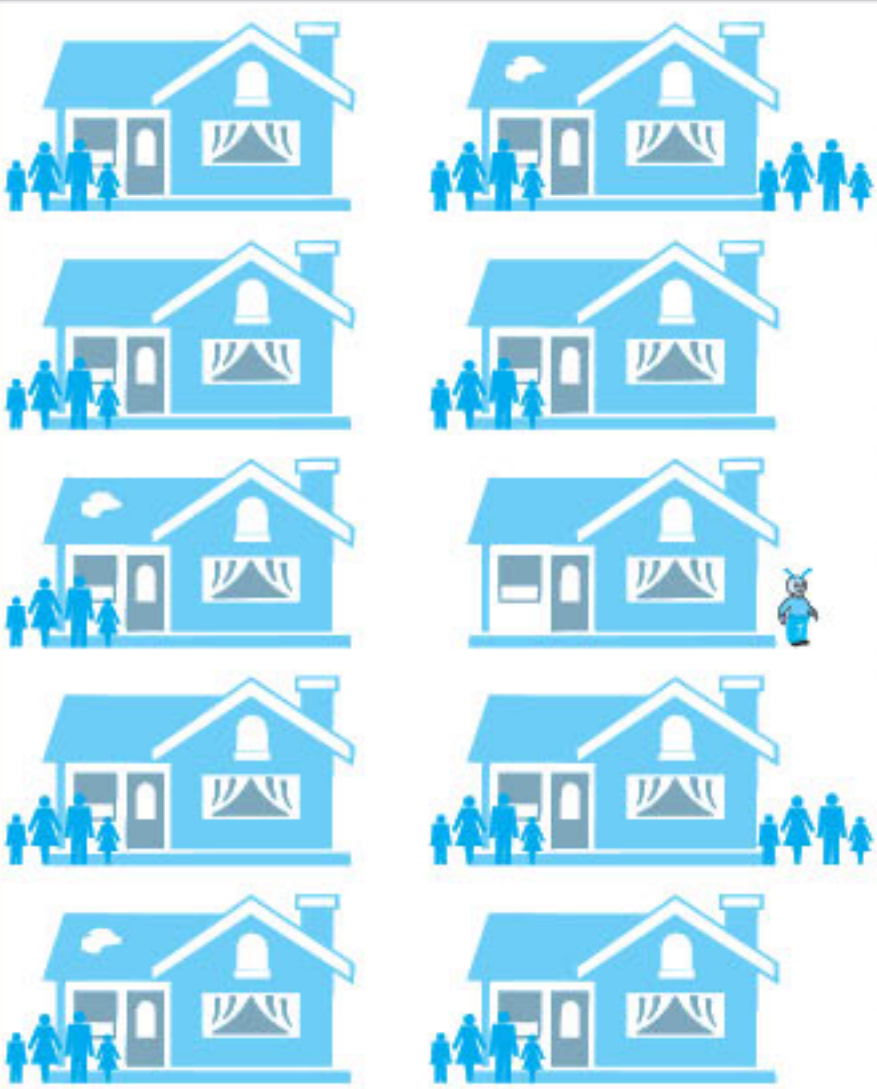
Units of Analysis	Sample Statements
Individuals 	60% of the sample are women 10% of the sample are wearing an eye patch 10% of the sample have pigtails
Families 	20% of the families have a single parent 50% of the families have two children 20% of the famillies have no children The mean number of children per family is 1.3
Households 	20% of the households are occupied by more than one family 30% of the households have holes in their roofs 10% of the households are occupied by aliens Notice also that 33% of the families live in multiple-family households with family as the unit of analysis

FIGURE 4-4

Illustrations of Units of Analysis. Units of analysis in social research can be individuals, groups, or even nonhuman entities.

implies a set of all objects of the same class: all books, all novels, all biographies, all introductory sociology textbooks, all cookbooks, all press conferences. In a study using books as the units of analysis, an individual book might be characterized by its size, weight, length, price, content, number of pictures, number sold, or description of the author. Then the population of all books or of a particular kind of book could be analyzed for the purpose of description or explanation: what kinds of books sell best and why, for example.

Similarly, a social researcher could analyze whether paintings by Russian, Chinese, or U.S. artists showed the greatest degree of working-class consciousness, taking paintings as the units of analysis and describing each, in part, by the nationality of its creator. Or you might examine a newspaper's editorials regarding a local university, for the purpose of describing, or perhaps explaining, changes in the newspaper's editorial position on the university over time. In this example, individual editorials would be the units of analysis.

Social interactions form another class of social artifacts suitable for social research. For example, we might characterize weddings as racially or religiously mixed or not, as religious or secular in ceremony, as resulting in divorce or not, or by descriptions of one or both of the marriage partners (such as "previously married," "Oakland Raider fan," "wanted by the FBI"). When a researcher reports that weddings between partners of different religions are more likely to be performed by secular authorities than those between partners of the same religion are, the weddings are the units of analysis, not the individuals involved.

Other social interactions that might be units of analysis are friendship choices, court cases, traffic accidents, divorces, fistfights, ship launchings, airline hijackings, race riots, final exams, student demonstrations, and congressional hearings. Congressional hearings, for instance, could be characterized by whether or not they occurred during an election campaign, whether the committee chairs were running for a higher office, whether they had received campaigns contributions from interested parties, and so on. Notice that even if we characterized and compared the hearings in terms of the committee chairs, the hearings themselves—not the individual chairpersons—would be our units of analysis.

Units of Analysis in Review

The examples in this section should suggest the nearly infinite variety of possible units of analysis in social research. Although individual human beings are typical objects of study, many research questions can be answered more appropriately through the examination of other units of analysis. Indeed, social researchers can study just about anything that bears on social life.

Moreover, the types of units of analysis named in this section do not begin to exhaust the possibilities. Morris Rosenberg (1968: 234–48), for example, speaks of individual, group, organizational, institutional, spatial, cultural, and societal units of analysis. John and Lyn Lofland (1995: 103–13) speak of practices, episodes, encounters, roles, relationships, groups, organizations, settlements, social worlds, lifestyles, and subcultures as suitable units of study. The important thing here is to grasp the logic of units of analysis. Once you do, the possibilities for fruitful research are limited only by your imagination.

Categorizing possible units of analysis might make the concept seem more complicated than it needs to be. What you call a given unit of analysis—a group, a formal organization, or a social artifact—is irrelevant. The key is to be clear about what your unit of analysis is. When you embark on a research project, you must decide whether you're studying marriages or marriage partners, crimes or criminals, corporations or corporate executives. Otherwise, you run the risk of drawing invalid conclusions because your assertions about one unit of analysis are actually based on the examination of another. We'll see an example of this issue in the next section as we look at the ecological fallacy.

Faulty Reasoning about Units of Analysis: The Ecological Fallacy and Reductionism

At this point, it's appropriate to introduce two types of faulty reasoning that you should be aware of: the ecological fallacy and reductionism. Each

represents a potential pitfall regarding units of analysis, and either can occur in doing research and drawing conclusions from the results.

The Ecological Fallacy

In this context, "ecological" refers to groups or sets or systems: something larger than individuals. The **ecological fallacy** is the assumption that something learned about an ecological unit says something about the individuals making up that unit. Let's consider a hypothetical illustration of this fallacy.

Suppose we're interested in learning something about the nature of electoral support received by a female political candidate in a recent citywide election. Let's assume we have the vote tally for each precinct so we can tell which precincts gave her the greatest support and which the least. Assume also that we have census data describing some characteristics of these precincts. Our analysis of such data might show that precincts with relatively young voters gave the female candidate a greater proportion of their votes than did precincts with older voters. We might be tempted to conclude from these findings that younger voters are more likely to vote for female candidates than older voters are—in other words, that age affects support for the woman. In reaching such a conclusion, we run the risk of committing the ecological fallacy, because it may have been the older voters in those "young" precincts who voted for the woman. Our problem is that we have examined *precincts* as our units of analysis but wish to draw conclusions about *voters*.

The same problem would arise if we discovered that crime rates were higher in cities having large African American populations than in those with few African Americans. We would not know if the crimes were actually committed by African Americans. Or if we found suicide rates higher in Protestant countries than in Catholic ones, we still could not know for sure that more Protestants than Catholics committed suicide.

In spite of these hazards, social researchers often have little choice but to address a particular research question through an ecological analysis. Perhaps the most appropriate data are simply not available. For example, the precinct vote tallies and the precinct characteristics mentioned in our initial example may be easy to obtain, but we may not have the resources to conduct a postelection survey of individual voters. In such cases, we may reach a tentative conclusion, recognizing and noting the risk of an ecological fallacy.

Although you should be careful not to commit the ecological fallacy, don't let these warnings lead you into committing what we might call the individualistic fallacy. Some people who approach social research for the first time have trouble reconciling general patterns of attitudes and actions with individual exceptions. But generalizations and probabilistic statements are not invalidated by individual exceptions. Your knowing a rich Democrat, for example, doesn't deny the fact that most rich people vote Republican—as a general pattern. Similarly, if you know someone who has gotten rich without any formal education, that doesn't deny the general pattern of higher education relating to higher income.

The ecological fallacy deals with something else altogether—confusing units of analysis in such a way that we draw conclusions about individuals solely from the observation of groups. Although the patterns observed between variables at the level of groups may be genuine, the danger lies in reasoning from the observed attributes of groups to the attributes of the individuals who made up those groups, even though we have not actually observed individuals.

ecological fallacy Erroneously drawing conclusions about individuals solely from the observation of groups.

Reductionism

A second type of potentially faulty reasoning related to units of analysis is reductionism. **Reductionism** involves attempts to explain a particular phenomenon in terms of limited and/or lower-order concepts. The reductionist explanation is not altogether wrong; it is simply too limited. Thus, you

might attempt to predict this year's winners and losers in the National Basketball Association by focusing on the abilities of the individual players on each team. This is certainly not stupid or irrelevant, but the success or failure of *teams* involves more than just the individuals in them; it involves coaching, teamwork, strategies, finances, facilities, fan loyalty, and so forth. To understand why some teams do better than others, you would make "team" the unit of analysis, and the quality of players would be one variable you would probably want to use in describing and classifying the teams.

Further, different academic disciplines approach the same phenomenon quite differently. Sociologists tend to consider sociological variables (such as values, norms, and roles), economists ponder economic variables (such as supply and demand and marginal value), and psychologists examine psychological variables (such as personality types and traumas). Explaining all or most human behavior in terms of economic factors is called economic reductionism, explaining it in terms of psychological factors is called psychological reductionism, and so forth. Notice how this issue relates to the discussion of theoretical paradigms in Chapter 2.

For many social scientists, the field of **sociobiology** is a prime example of reductionism, suggesting that all social phenomena can be explained in terms of biological factors. Thus, for example, Edward O. Wilson (1975) sought to explain altruistic behavior in human beings in terms of genetic makeup. In his neo-Darwinian view, Wilson suggests that humans have evolved in such a way that individuals sometimes need to sacrifice themselves for the benefit of the whole species. Some people might explain such sacrifice in terms of ideals or warm feelings between humans. However, genes are the essential unit in Wilson's paradigm, producing his famous dictum that human beings are "only DNA's way of making more DNA."

Reductionism of any type tends to suggest that particular units of analysis or variables are more relevant than others. Suppose we ask what caused the American Revolution. Was it a shared commitment to the value of individual liberty? The economic plight of the colonies in relation to Britain? The megalomania of the founders? As soon as we inquire about *the* single cause, we run the risk of reductionism. If we were to regard shared values as the cause of the American Revolution, our unit of analysis would be the individual colonist. An economist, though, might choose the 13 colonies as units of analysis and examine the economic organizations and conditions of each. A psychologist might choose individual leaders as the units of analysis for purposes of examining their personalities. Of course, there's nothing wrong in choosing these units of analysis as part of an explanation of the American Revolution, but I think you can see how each alone would not produce a complete answer.

Like the ecological fallacy, reductionism can occur when we use inappropriate units of analysis. The appropriate unit of analysis for a given research question, however, is not always clear. Social researchers, especially across disciplinary boundaries, often debate this issue.

The Time Dimension

So far in this chapter, we've regarded research design as a process for deciding what aspects we'll observe, of whom, and for what purpose. Now we must consider a set of time-related options that cuts across each of these earlier considerations. We can choose to make observations more or less at one time or over a long period.

Time plays many roles in the design and execution of research, quite aside from the time it takes to do research. Earlier we noted that the time sequence of events and situations is critical to

reductionism A fault of some researchers: a strict limitation (reduction) of the kinds of concepts to be considered relevant to the phenomenon under study.

sociobiology A paradigm based in the view that social behavior can be explained solely in terms of genetic characteristics and behavior.

determining causation (a point we'll return to in Part 4). Time also affects the generalizability of research findings. Do the descriptions and explanations resulting from a particular study accurately represent the situation of ten years ago, ten years from now, or only the present? Researchers have two principal options available to deal with the issue of time in the design of their research: cross-sectional studies and longitudinal studies.

Cross-Sectional Studies

A **cross-sectional study** involves observations of a sample, or cross section, of a population or phenomenon that are made at one point in time. Exploratory and descriptive studies are often cross-sectional. A single U.S. Census, for instance, is a study aimed at describing the U.S. population at a given time.

Many explanatory studies are also cross-sectional. A researcher conducting a large-scale national survey to examine the sources of racial and religious prejudice would, in all likelihood, be dealing with a single time frame—taking a snapshot, so to speak, of the sources of prejudice at a particular point in history.

Explanatory cross-sectional studies have an inherent problem. Although their conclusions are based on observations made at only one time, typically they aim at understanding causal processes that occur over time. This problem is somewhat akin to that of determining the speed of a moving object on the basis of a high-speed, still photograph that freezes the movement of the object.

Yanjie Bian, for example, conducted a survey of workers in Tianjin, China, for the purpose of studying stratification in contemporary, urban Chinese society. In undertaking the survey in 1988, however, he was conscious of the important changes brought about by a series of national campaigns, such as the Great Proletarian Cultural Revolution, dating from the Chinese Revolution in 1949 (which brought the Chinese Communists into power) and continuing into the present.

> These campaigns altered political atmospheres and affected people's work and nonwork activities. Because of these campaigns, it is difficult to draw conclusions from a cross-sectional social survey, such as the one presented in this book, about general patterns of Chinese workplaces and their effects on workers. Such conclusions may be limited to one period of time and are subject to further tests based on data collected at other times.
>
> *(1994: 19)*

The problem of generalizations about social life from a "snapshot" is one this book repeatedly addresses. One solution is suggested by Bian's final comment—about data collected "at other times": Social research often involves revisiting phenomena and building on the results of earlier research.

cross-sectional study A study based on observations representing a single point in time.

longitudinal study A study design involving the collection of data at different points in time.

Longitudinal Studies

In contrast to cross-sectional studies, a **longitudinal study** is designed to permit observations of the same phenomenon over an extended period. For example, a researcher can participate in and observe the activities of a UFO cult from its inception to its demise. Other longitudinal studies use records or artifacts to study changes over time. In analyses of newspaper editorials or Supreme Court decisions over time, for example, the studies are longitudinal whether the researcher's actual observations and analyses were made at one time or over the course of the actual events under study.

Many field research projects, involving direct observation and perhaps in-depth interviews, are naturally longitudinal. Thus, for example, when Ramona Asher and Gary Fine (1991) studied the life experiences of the wives of alcoholic men, they were in a position to examine the evolution of troubled marital relationships over time, sometimes

even including the reactions of the subjects to the research itself.

In the classic study *When Prophecy Fails* (1956), Leon Festinger, Henry Reicker, and Stanley Schachter were specifically interested in learning what happened to a flying saucer cult when their predictions of an alien encounter failed to come true. Would the cult members close down the group, or would they become all the more committed to their beliefs? A longitudinal study was required to provide an answer. (They redoubled their efforts to get new members.)

Longitudinal studies can be more difficult for quantitative studies such as large-scale surveys. Nonetheless, they are often the best way to study changes over time. There are three special types of longitudinal studies that you should know about: trend studies, cohort studies, and panel studies.

Trend Studies

A **trend study** is a type of longitudinal study that examines changes within a population over time. A simple example is a comparison of U.S. Censuses over a period of decades, showing shifts in the makeup of the national population. A similar use of archival data was made by Michael Carpini and Scott Keeter (1991), who wanted to know whether contemporary U.S. citizens were better or more poorly informed about politics than citizens of an earlier generation were. To find out, they compared the results of several Gallup Polls conducted during the 1940s and 1950s with a 1989 survey that asked several of the same questions tapping political knowledge.

Overall, the analysis suggested that contemporary citizens were slightly better informed than earlier generations were. In 1989, 74 percent of the sample could name the vice president of the United States, compared with 67 percent in 1952. Substantially higher percentages of people in 1989 than in 1947 could explain presidential vetoes and congressional overrides of vetoes. On the other hand, more of the 1947 sample could identify their U.S. representative (38 percent) than the 1989 sample (29 percent) could.

An in-depth analysis, however, indicates that the slight increase in political knowledge resulted from the fact that the people in the 1989 sample were more highly educated than those from earlier samples were. When educational levels were taken into account, the researchers concluded that political knowledge has actually declined within specific educational groups.

Cohort Studies

In a **cohort study**, a researcher examines specific subpopulations, or *cohorts*, as they change over time. Typically, a cohort is an age group, such as people born during the 1950s, but it can also be some other time grouping, such as people born during the Vietnam War, people who got married in 1994, and so forth. An example of a cohort study would be a series of national surveys, conducted perhaps every 20 years, to study the attitudes of the cohort born during World War II toward U.S. involvement in global affairs. A sample of people 15–20 years old might be surveyed in 1960, another sample of those 35–40 years old in 1980, and another sample of those 55–60 years old in 2000. Although the specific set of people studied in each survey would differ, each sample would represent the cohort born between 1940 and 1945.

James Davis (1992) turned to a cohort analysis in an attempt to understand shifting political orientations during the 1970s and 1980s in the United States. Overall, he found a liberal trend on issues

trend study A type of longitudinal study in which a given characteristic of some population is monitored over time. An example would be the series of Gallup Polls showing the electorate's preferences for political candidates over the course of a campaign, even though different samples were interviewed at each point.

cohort study A study in which some specific subpopulation, or cohort, is studied over time, although data may be collected from different members in each set of observations. For example, a study of the occupational history of the class of 1970 in which questionnaires were sent every five years would be a cohort study.

TABLE 4-1

Age and Political Liberalism

Survey dates	*1972 to 1974*	*1977 to 1980*	*1982 to 1984*	*1987 to 1989*
Age of cohort	20–24	25–29	30–34	35–39
Percent who would let the Communist speak	72%	68%	73%	73%

such as race, gender, religion, politics, crime, and free speech. But did this trend represent people in general getting a bit more liberal, or did it merely reflect liberal younger generations replacing the conservative older ones?

To answer this question, Davis examined national surveys (from the General Social Survey, of which he is a founder) conducted in four time periods, five years apart. In each survey, he grouped the respondents into age groups, also five years apart. This strategy allowed him to compare different age groups at any given point in time as well as follow the political development of each age group over time.

One of the questions he examined was whether a person who admitted to being a Communist should be allowed to speak in the respondents' communities. Consistently, the younger respondents in each time period were more willing to let the Communist speak than the older ones were. Among those aged 20–40 in the first set of the survey, for example, 72 percent took this liberal position, contrasted with 27 percent among respondents 80 and older. What Davis found when he examined the youngest cohort over time is shown in Table 4-1. This pattern of a slight, conservative shift in the 1970s, followed by a liberal rebound in the 1980s, typifies the several cohorts Davis analyzed (J. Davis 1992: 269).

In another study, Eric Plutzer and Michael Berkman (2005) used a cohort design to completely reverse a prior conclusion regarding aging and support for education. Logically, as people grow well beyond the child-rearing years, we might expect them to reduce their commitment to educational funding. Moreover, cross-sectional data support that expectation. The researchers present several data sets showing those over 65 voicing less support for education funding than those under 65 did.

Such simplistic analyses, however, leave out an important variable: increasing support for educational funding in U.S. society over time in general. The researchers add to this the concept of "generational replacement," meaning that the older respondents in a survey grew up during a time when there was less support for education in general, whereas the younger respondents grew up during a time of greater overall support.

A cohort analysis allowed the researchers to determine what happened to the attitudes of specific cohorts over time. Here, for example, are the percentages of Americans born during the 1940s who felt educational spending was too low, when members of that cohort were interviewed over time (Plutzer and Berkman, 2000: 76):

Year Interviewed	*Percent Who Say Educational Funding Is Too Low*
1970s	58
1980s	66
1990s	74
2000s	79

As these data indicate, those who were born during the 1940s have steadily increased their support for educational funding as they have passed through and beyond the child-rearing years.

panel study A type of longitudinal study, in which data are collected from the same set of people (the sample or panel) at several points in time.

Panel Studies

Though similar to trend and cohort studies, a **panel study** examines the same set of people each

time. For example, we could interview the same sample of voters every month during an election campaign, asking for whom they intended to vote. Though such a study would allow us to analyze overall trends in voter preferences for different candidates, it would also show the precise patterns of persistence and change in intentions. For example, a trend study that showed that Candidates A and B each had exactly half of the voters on September 1 and on October 1 as well could indicate that none of the electorate had changed voting plans, that all of the voters had changed their intentions, or something in-between. A panel study would eliminate this confusion by showing what kinds of voters switched from A to B and what kinds switched from B to A, as well as other facts.

Joseph Veroff, Shirley Hatchett, and Elizabeth Douvan (1992) wanted to learn about marital adjustment among newlyweds, specifically regarding differences between white and African American couples. To get subjects for study, they selected a sample of couples who applied for marriage licenses in Wayne County, Michigan, April through June 1986.

Concerned about the possible impact their research might have on the couples' marital adjustment, the researchers divided their sample in half at random: an *experimental* group and a *control* group (concepts we'll explore further in Chapter 8). Couples in the former group were intensively interviewed over a four-year period, whereas the latter group was contacted only briefly each year.

By studying the same couples over time, the researchers could follow the specific problems that arose and the way the couples dealt with them. As a by-product of their research, they found that those studied the most intensely seemed to achieve a somewhat better marital adjustment. The researchers felt that the interviews could have forced couples to discuss matters they might have otherwise buried.

Comparing the Three Types of Longitudinal Studies

To reinforce the distinctions among trend, cohort, and panel studies, let's contrast the three study designs in terms of the same variable: *religious affiliation*. A trend study might look at shifts in U.S. religious affiliations over time, as the Gallup Poll does on a regular basis. A cohort study might follow shifts in religious affiliations among "the Depression generation," specifically, say, people who were 20 to 30 years old in 1932. We could study a sample of people 30–40 years old in 1942, a new sample of people aged 40–50 in 1952, and so forth. A panel study could start with a sample of the whole population or of some special subset and study those specific individuals over time. Notice that only the panel study would give a full picture of the shifts among the various categories of affiliations, including "none." Cohort and trend studies would uncover only net changes.

Longitudinal studies have an obvious advantage over cross-sectional ones in providing information describing processes over time. But this advantage often comes at a heavy cost in both time and money, especially in a large-scale survey. Observations may have to be made at the time events are occurring, and the method of observation may require many research workers.

Panel studies, which offer the most comprehensive data on changes over time, face a special problem: panel attrition. Some of the respondents studied in the first wave of the survey might not participate in later waves. (This is comparable to the problem of experimental mortality discussed in Chapter 8.) The danger is that those who drop out of the study may be atypical, thereby distorting the results of the study. Thus, when Carol Aneshensel and her colleagues conducted a panel study of adolescent girls (comparing Latinas and non-Latinas), they looked for and found differences in characteristics of survey dropouts among Latinas born in the United States and those born in Mexico. These differences needed to be taken into account to avoid misleading conclusions about differences between Latinas and non-Latinas (Aneshensel et al. 1989).

Approximating Longitudinal Studies

Longitudinal studies do not always provide a feasible or practical means of studying processes that take place over time. Fortunately, researchers often can draw approximate conclusions about such

processes even when only cross-sectional data are available. Here are some ways to do that.

Sometimes cross-sectional data imply processes over time on the basis of simple logic. For example, in the study of student drug use conducted at the University of Hawaii (Chapter 2), students were asked to report whether they had ever tried each of several illegal drugs. The study found that some students had tried both marijuana and LSD, some had tried only one, and others had tried neither. Because these data were collected at one time, and because some students presumably would experiment with drugs later on, it would appear that such a study could not tell whether students were more likely to try marijuana or LSD first.

A closer examination of the data showed, however, that although some students reported having tried marijuana but not LSD, there were no students in the study who had tried only LSD. From this finding it was inferred—as common sense suggested—that marijuana use preceded LSD use. If the process of drug experimentation occurred in the opposite time order, then a study at a given time should have found some students who had tried LSD but not marijuana, and it should have found no students who had tried only marijuana.

Researchers can also make logical inferences whenever the time order of variables is clear. If we discovered in a cross-sectional study of college students that those educated in private high schools received better college grades than those educated in public high schools did, we would conclude that the type of high school attended affected college grades, not the other way around. Thus, even though we made our observations at only one time, we would feel justified in drawing conclusions about processes taking place across time.

Very often, age differences discovered in a cross-sectional study form the basis for inferring processes across time. Suppose you're interested in the pattern of worsening health over the course of the typical life cycle. You might study the results of annual checkups in a large hospital. You could group health records according to the ages of those examined and rate each age group in terms of several health conditions—sight, hearing, blood pressure, and so forth. By reading across the age-group ratings for each health condition, you would have something approximating the health history of individuals. Thus, you might conclude that the average person develops vision problems before hearing problems. You would need to be cautious in this assumption, however, because the differences might reflect societywide trends. Perhaps improved hearing examinations instituted in the schools had affected only the young people in your study.

Asking people to recall their pasts is another common way of approximating observations over time. Researchers use that method when they ask people where they were born or when they graduated from high school or whom they voted for in 1988. Qualitative researchers often conduct in-depth "life history" interviews. For example, C. Lynn Carr (1998) used this technique in a study of "tomboyism." Her respondents, aged 25–40, were asked to reconstruct aspects of their lives from childhood on, including experiences of identifying themselves as tomboys.

The danger in this technique is evident. Sometimes people have faulty memories; sometimes they lie. When people are asked in postelection polls whom they voted for, the results inevitably show more people voting for the winner than actually did so on election day. As part of a series of in-depth interviews, such a report can be validated in the context of other reported details; however, results based on a single question in a survey must be regarded with caution.

This discussion of the ways that time figures into social research suggests several questions you should confront in your own research projects. In designing any study, be sure to look at both the explicit and implicit assumptions you're making about time. Are you interested in describing some process that occurs over time, or are you simply going to describe what exists now? If you want to describe a process occurring over time, will you be able to make observations at different points in the process, or will you have to approximate such observations by drawing logical inferences from what you can observe now? If you opt for a longitudinal design, which method best serves your research purposes?

Examples of Research Strategies

As the preceding discussions have implied, social research follows many paths. The following short excerpts further illustrate this point. As you read each excerpt, note both the content of each study and the method used to study the chosen topic. Does the study seem to be exploring, describing, or explaining (or some combination of these)? What are the sources of data in each study? Can you identify the unit of analysis? Is the dimension of time relevant? If so, how will it be handled?

- This case study of unobtrusive mobilizing by Southern California Rape Crisis Center uses archival, observational, and interview data to explore how a feminist organization worked to change police, schools, prosecutors, and some state and national organizations from 1974 to 1994. (Schmitt and Martin 1999: 364)
- Using life history narratives, the present study investigates processes of agency and consciousness among 14 women who identified themselves as tomboys. (Carr 1998: 528)
- By drawing on interviews with activists in the former Estonian Soviet Socialist Republic, we specify the conditions by which accommodative and oppositional subcultures exist and are successfully transformed into social movements. (Johnston and Snow 1998: 473)
- This paper presents the results of an ethnographic study of an AIDS service organization located in a small city. It is based on a combination of participant observation, interviews with participants, and review of organizational records. (Kilburn 1998: 89)
- Using interviews obtained during fieldwork in Palestine in 1992, 1993, and 1994, and employing historical and archival records, I argue that Palestinian feminist discourses were shaped and influenced by the sociopolitical context in which Palestinian women acted and with which they interacted. (Abdulhadi 1998: 649)
- This article reports on women's experiences of breastfeeding in public as revealed through in-depth interviews with 51 women. (Stearns 1999: 308)
- Using interview and observational field data, I demonstrate how a system of temporary employment in a participative workplace both exploited and shaped entry-level workers' aspirations and occupational goals. (V. Smith 1998: 411)
- I collected data [on White Separatist rhetoric] from several media of public discourse, including periodicals, books, pamphlets, transcripts from radio and television talk shows, and newspaper and magazine accounts. (Berbrier 1998: 435)
- In the analysis that follows, racial and gender inequality in employment and retirement will be analyzed, using a national sample of persons who began receiving Social Security Old Age benefits in 1980–81. (Hogan and Perrucci 1998: 528)
- Drawing from interviews with female crack dealers, this paper explores the techniques they use to avoid arrest. (Jacobs and Miller 1998: 550)

How to Design a Research Project

You've now seen some of the options available to social researchers in designing projects. I know there are a lot of components, and the relationships among them may not be totally clear, so here's a way of pulling them together. Let's assume you were to undertake research. Where would you start? Then, where would you go?

Although research design occurs at the beginning of a research project, it involves all the steps of the subsequent project. This discussion, then, provides both guidance on how to start a research project and an overview of the topics that follow in later chapters of this book.

Figure 4-5 presents a schematic view of the social research process. I present this view reluctantly, because it may suggest more of a step-by-step order to research than actual practice bears out. Nonetheless, this idealized overview of the process

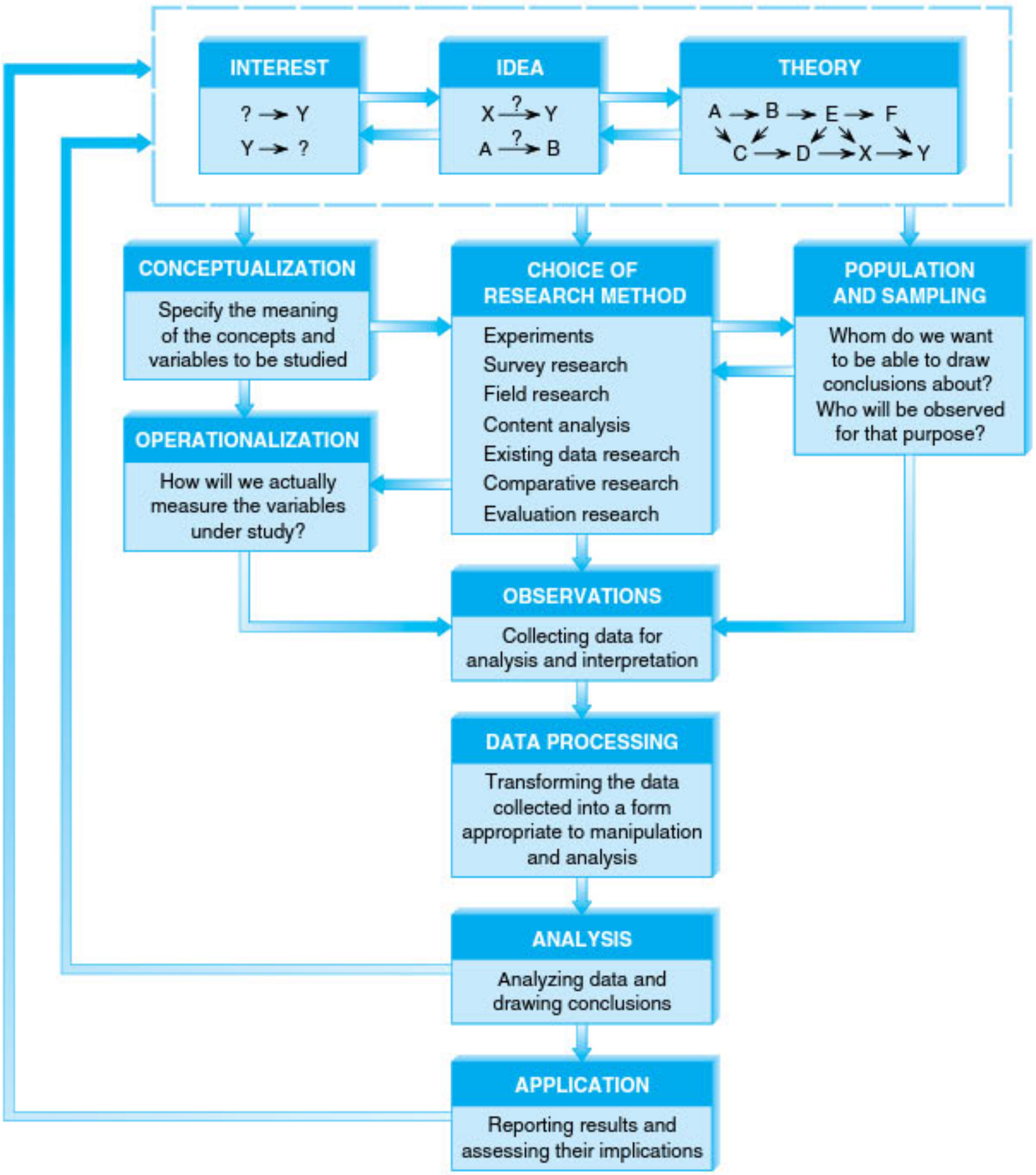


FIGURE 4-5

The Research Process. Here are some of the key elements that we'll be examining throughout this book: the pieces that make up the whole of social research.

provides a context for the specific details of particular components of social research. Essentially, it is another and more detailed picture of the scientific process presented in Chapter 2.

At the top of the diagram are interests, ideas, and theories, the possible beginning points for a line of research. The letters (A, B, X, Y, and so forth) represent variables or concepts such as prejudice or alienation. Thus, you might have a general interest in finding out what causes some people to be more prejudiced than others, or you might want to know some of the consequences of alienation. Alternatively, your inquiry might begin with a specific idea about the way things are. For example, you might have the idea that working on an assembly line causes alienation. The question marks in the diagram indicate that you aren't sure things are the way you suspect they are—that's why you're doing the research. Notice that a theory is represented as a set of complex relationships among several variables.

The double arrows between "interest," "idea," and "theory" suggest that there is often a movement back and forth across these several possible beginnings. An initial interest may lead to the formulation of an idea, which may be fit into a larger theory, and the theory may produce new ideas and create new interests.

Any or all of these three may suggest the need for empirical research. The purpose of such research can be to explore an interest, test a specific idea, or validate a complex theory. Whatever the purpose, the researcher needs to make a variety of decisions, as indicated in the remainder of the diagram.

To make this discussion more concrete, let's take a specific research example. Suppose you're concerned with the issue of abortion and have a special interest in learning why some college students support abortion rights and others oppose them. Going a step further, let's say you've formed the impression that students in the humanities and social sciences seem generally more inclined to support the idea of abortion rights than those in the natural sciences do. (That kind of thinking often leads people to design and conduct social research.)

So, where do you start? You have an idea you want to pursue, one that involves abortion attitudes and choice of college major. In terms of the options we've discussed in this chapter, you probably have both descriptive and explanatory interests, but you might decide you only want to explore the issue. You might wonder what sorts of attitudes students with different majors have about abortion (exploratory), what percentage of the student body supports a woman's right to an abortion (descriptive), or what causes some to support it and others to oppose it (explanation). The units of analysis in this case would individuals: college students. But we're jumping the gun. As you can see, even before we've "started," we've started. The reciprocal processes described in Figure 4-5 begin even before you've made a commitment to a project. Let's look more formally at the various steps, then, keeping this reciprocal motion in mind.

Getting Started

At the outset of your project, then, your aim would probably be exploratory. At this point, you might choose among several possible activities in pursuing your interest in student attitudes about abortion rights. To begin with, you might want to read something about the issue. If you have a hunch that attitudes are somehow related to college major, you might find out what other researchers may have written about that. Appendix A of this book will help you make use of your college library. In addition, you would probably talk to some people who support abortion rights and some who don't. You might attend meetings of abortion-related groups. All these activities could help prepare you to handle the various decisions of research design we're about to examine.

Before designing your study, you must define the purpose of your project. What kind of study will you undertake—exploratory, descriptive, explanatory? Do you plan to write a research paper to satisfy a course or thesis requirement? Is your purpose to gain information that will support you in arguing for or against abortion rights? Do you want to write an article for the campus newspaper or an

academic journal? In reviewing the previous research literature regarding abortion rights, you should note the design decisions other researchers have made, always asking whether the same decisions would satisfy your purpose.

Usually, your purpose for undertaking research can be expressed as a report. A good first step in designing your project is to outline such a report (see Chapter 17 for help on this). Although your final report may not look much like your initial image of it, this exercise will help you figure out which research designs are most appropriate. During this step, clearly describe the kinds of statements you want to make when the research is complete. Here are some examples of such statements: "Students frequently mentioned abortion rights in the context of discussing social issues that concerned them personally." "X percent of State U. students favor a woman's right to choose an abortion." "Engineers are (more/less) likely than sociologists to favor abortion rights."

Conceptualization

Once you have a well-defined purpose and a clear description of the kinds of outcomes you want to achieve, you can proceed to the next step in the design of your study—conceptualization. We often talk pretty casually about social science concepts such as prejudice, alienation, religiosity, and liberalism, but it's necessary to clarify what we mean by these concepts, in order to draw meaningful conclusions about them. Chapter 5 examines this process of conceptualization in depth. For now, let's see what it might involve in the case of our hypothetical example.

If you're going to study how college students feel about abortion and why, the first thing you'll have to specify is what you mean by "the right to an abortion." Because support for abortion probably varies according to the circumstances, you'll want to pay attention to the different conditions under which people might approve or disapprove of abortion: for example, when the woman's life is in danger, in the case of rape or incest, or simply as a matter of personal choice.

Similarly, you'll need to specify exact meanings for all the other concepts you plan to study. If you want to study the relationship of opinion about abortion to college major, you'll have to decide whether you want to consider only officially declared majors or to include students' intentions as well. What will you do with those who have no major?

In surveys and experiments, you need to specify such concepts in advance. In less tightly structured research, such as open-ended interviews, an important part of the research may involve the discovery of different dimensions, aspects, or nuances of concepts. In such cases, the research itself may uncover and report aspects of social life that were not evident at the outset of the project.

Choice of Research Method

As we'll discuss in Part 3, each research method has its strengths and weaknesses, and certain concepts are more appropriately studied through some methods than through others. In our study of attitudes toward abortion rights, a survey might be the most appropriate method: either interviewing students or asking them to fill out a questionnaire. Surveys are particularly well suited to the study of public opinion. This is not to say that you couldn't make good use of the other methods presented in Part 3. For example, you might use the method of content analysis to examine letters to the editor and analyze the different images of abortion that letter writers have. Field research would provide an avenue to understanding how people interact with one another regarding the issue of abortion, how they discuss it, and how they change their minds. Other research methods introduced in Part 3 could also be used in studying this topic. Usually, the best study design uses more than one research method, taking advantage of their different strengths. If you look back at the brief examples of actual studies at the end of the preceding section, you'll see several instances where the researchers used many methods in a single study.

Operationalization

Once you've specified the concepts to be studied and chosen a research method, the next step is operationalization, or deciding on your measurement techniques (discussed further in Chapters 5 and 6). The meaning of variables in a study is determined in part by how they are measured. Part of the task here is deciding how the desired data will be collected: direct observation, review of official documents, a questionnaire, or some other technique.

If you decided to use a survey to study attitudes toward abortion rights, part of operationalization is determining the wording of questionnaire items. For example, you might operationalize your main variable by asking respondents whether they would approve of a woman's right to have an abortion under each of the conditions you've conceptualized: in the case of rape or incest, if her life were threatened by the pregnancy, and so forth. You'd design the questionnaire so that it asked respondents to express approval or disapproval for each situation. Similarly, you would specify exactly how respondents would indicate their college major, as well as what choices to provide those who have not declared a major.

Population and Sampling

In addition to refining concepts and measurements, you must decide whom or what to study. The *population* for a study is that group (usually of people) about whom we want to draw conclusions. We're almost never able to study all the members of the population that interests us, however, and we can never make every possible observation of them. In every case, then, we select a *sample* from among the data that might be collected and studied. The sampling of information, of course, occurs in everyday life and often produces biased observations. (Recall the discussion of "selective observation" in Chapter 1.) Social researchers are more deliberate in their sampling of what will be observed.

Chapter 7 describes methods for selecting samples that adequately reflect the whole population that interests us. Notice in Figure 4-5 that decisions about population and sampling are related to decisions about the research method to be used. Whereas probability sampling techniques would be relevant to a large-scale survey or a content analysis, a field researcher might need to select only those informants who will yield a balanced picture of the situation under study, and an experimenter might assign subjects to experimental and control groups in a manner that creates comparability.

In your hypothetical study of abortion attitudes, the relevant population would be the student population of your college. As you'll discover in Chapter 7, however, selecting a sample will require you to get more specific than that. Will you include part-time as well as full-time students? Only degree candidates or everyone? International students as well as U.S. citizens? Undergraduates, graduate students, or both? There are many such questions—each of which must be answered in terms of your research purpose. If your purpose is to predict how students would vote in a local referendum on abortion, you might want to limit your population to those eligible and likely to vote.

Observations

Having decided what to study among whom by what method, you're now ready to make observations—to collect empirical data. The chapters of Part 3, which describe the various research methods, give the different observation techniques appropriate to each.

To conduct a survey on abortion, you might want to print questionnaires and mail them to a sample selected from the student body. Alternatively, you could arrange to have a team of interviewers conduct the survey over the telephone. The relative advantages and disadvantages of these and other possibilities are discussed in Chapter 9.

Data Processing

Depending on the research method chosen, you'll have amassed a volume of observations in a form that probably isn't immediately interpretable. If

you've spent a month observing a street-corner gang firsthand, you'll now have enough field notes to fill a book. In a historical study of ethnic diversity at your school, you may have amassed volumes of official documents, interviews with administrators and others, and so forth. Chapters 13 and 14 describe some of the ways social scientific data are processed or transformed for qualitative or quantitative analysis.

In the case of a survey, the "raw" observations are typically in the form of questionnaires with boxes checked, answers written in spaces, and the like. The data-processing phase of a survey typically involves the classification (coding) of written-in answers and the transfer of all information to a computer.

Analysis

Once the collected data are in a suitable form, you're ready to interpret them for the purpose of drawing conclusions that reflect the interests, ideas, and theories that initiated the inquiry. Chapters 13 and 14 describe a few of the many options available to you in analyzing data. In Figure 4-5, notice that the results of your analyses feed back into your initial interests, ideas, and theories. Often this feedback represents the beginning of another cycle of inquiry.

In the survey of student attitudes about abortion rights, the analysis phase would pursue both descriptive and explanatory aims. You might begin by calculating the percentages of students who favored or opposed each of the several different versions of abortion rights. Taken together, these several percentages would provide a good picture of student opinion on the issue.

Moving beyond simple description, you might describe the opinions of subsets of the student body, such as different college majors. Provided that your design called for trapping other information about respondents, you could also look at men versus women; freshmen, sophomores, juniors, seniors, and graduate students; or other categories that you've included. The description of subgroups could then lead you into an explanatory analysis.

Application

The final stage of the research process involves the uses made of the research you've conducted and the conclusions you've reached. To start, you'll probably want to communicate your findings so that others will know what you've learned. It may be appropriate to prepare—and even publish—a written report. Perhaps you'll make oral presentations, such as papers delivered to professional and scientific meetings. Other students would also be interested in hearing what you've learned about them.

You may want to go beyond simply reporting what you've learned to discussing the implications of your findings. Do they say anything about actions that might be taken in support of policy goals? Both the proponents and the opponents of abortion rights would be interested.

Finally, be sure to consider what your research suggests in regard to further research on your subject. What mistakes should be corrected in future studies? What avenues—opened up slightly in your study—should be pursued further?

Research Design in Review

As this overview shows, research design involves a set of decisions regarding what topic is to be studied among what population with what research methods for what purpose. Although you'll want to consider many ways of studying a subject—and use your imagination as well as your knowledge of a variety of methods—research design is the process of focusing your perspective for the purposes of a particular study.

If you're doing a research project for one of your courses, many aspects of research design may be specified for you in advance, including the method (such as an experiment) or the topic (as in a course on a particular subject, such as prejudice). The following summary assumes that you're free to choose both your topic and your research strategy.

In designing a research project, you'll find it useful to begin by assessing three things: your interests, your abilities, and the available resources.

Each of these considerations will suggest a large number of possible studies.

Simulate the beginning of a somewhat conventional research project: Ask yourself what you're interested in understanding. Surely you have several questions about social behavior and attitudes. Why are some people politically liberal and others politically conservative? Why are some people more religious than others? Why do people join militia groups? Do colleges and universities still discriminate against minority faculty members? Why would a woman stay in an abusive relationship? Spend some time thinking about the kinds of questions that interest and concern you.

Once you have a few questions you'd be interested in answering for yourself, think about the kind of information needed to answer them. What research units of analysis would provide the most relevant information: college students, corporations, voters, cities, or corporations? This question will probably be inseparable in your thoughts from the question of research topics. Then ask which aspects of the units of analysis would provide the information you need in order to answer your research question.

Once you have some ideas about the kind of information relevant to your purpose, ask yourself how you might go about getting that information. Are the relevant data likely to be already available somewhere (say, in a government publication), or would you have to collect them yourself? If you think you would have to collect them, how would you go about doing it? Would you need to survey a large number of people or interview a few people in depth? Could you learn what you need to know by attending meetings of certain groups? Could you glean the data you need from books in the library?

As you answer these questions, you'll find yourself well into the process of research design. Keep in mind your own research abilities and the resources available to you. There's little point in designing a perfect study that you can't actually carry out. You may want to try a research method you haven't used before so you can learn from it, but be careful not to put yourself at too great a disadvantage.

Once you have a general idea of what you want to study and how, carefully review previous research in journals and books to see how other researchers have addressed the topic and what they have learned about it. Your review of the literature may lead you to revise your research design: Perhaps you'll decide to use a previous researcher's method or even replicate an earlier study. A standard procedure in the physical sciences, the independent replication of research projects is just as important in the social sciences, although social researchers tend to overlook that. Or, you might want to go beyond replication and study some aspect of the topic that you feel previous researchers have overlooked.

Here's another approach you might take. Suppose a topic has been studied previously using field research methods. Can you design an experiment that would test the findings those earlier researchers produced? Or, can you think of existing statistics that could be used to test their conclusions? Did a mass survey yield results that you'd like to explore in greater detail through on-the-spot observations and in-depth interviews? The use of several different research methods to test the same finding is sometimes called *triangulation*, and you should always keep it in mind as a valuable research strategy. Because each research method has particular strengths and weaknesses, there is always a danger that research findings will reflect, at least in part, the method of inquiry. In the best of all worlds, your own research design should bring more than one research method to bear on the topic.

The Research Proposal

Quite often, in the design of a research project, you'll have to lay out the details of your plan for someone else's review and/or approval. In the case of a course project, for example, your instructor might very well want to see a "proposal" before you set off to work. Later in your career, if you wanted to undertake a major project, you might need to obtain funding from a foundation or government agency, who would most definitely want a

detailed proposal that describes how you would spend their money. You may respond to a Request for Proposals (RFP), which both public and private agencies often circulate in search of someone to do research for them.

This chapter concludes with a brief discussion of how you might prepare a research proposal. This will give you one more overview of the whole research process that the rest of this book details.

Elements of a Research Proposal

Although some funding agencies (or your instructor, for that matter) may have specific requirements for the elements or structure of a research proposal, here are some basic elements you should include.

Problem or Objective

What exactly do you want to study? Why is it worth studying? Does the proposed study have practical significance? Does it contribute to the construction of social theories?

Literature Review

What have others said about this topic? What theories address it and what do they say? What previous research exists? Are there consistent findings, or do past studies disagree? Are there flaws in the body of existing research that you think you can remedy?

Chapter 17 has a lengthier discussion of this topic. You'll find that special skills involved in reading social science research reports requires special skills. If you need to undertake a review of the literature at this point in your course, you may want to skip ahead to Chapter 17. It will familiarize you with the different types of research literature, how to find what you want, and how to read it. There is a special discussion of how to use electronic resources online and how to avoid being misled by information on the Internet.

In part, your review of the literature will be shaped by the data-collection method(s) you intend to use in your study. Reviewing the designs of previous studies using that same technique can give you a head start in planning your own study. At the same time, you should focus your search on your research topic: regardless of the methods other researchers have used. So, if you're planning field research on, say, interracial marriages, you might gain some useful insights from the findings of surveys on the topic; further, past field research on interracial marriages could be invaluable in your designing a survey on the topic.

Because the literature review will appear early in your research proposal, you should write it with an eye to introducing the reader to the topic you will address, laying out in a logical manner what has already been learned on the topic by past researchers, then leading up to the holes or loose ends in our knowledge of the topic, which you propose to remedy. Or a little differently, your review of the literature may point to inconsistencies or disagreements to be found among the existing research findings. In that case, your proposed research will aim to resolve the ambiguities that plague us. I don't know about you, but I'm already excited about the research you're proposing to undertake.

Subjects for Study

Whom or what will you study in order to collect data? Identify the subjects in general, theoretical terms; in specific, more concrete terms, identify who is available for study and how you'll reach them. Will it be appropriate to select a sample? If so, how will you do that? If there is any possibility that your research will affect those you study, how will you insure that the research does not harm them?

Beyond these general questions, the specific research method you'll use will further specify the matter. If you're planning to undertake an experiment, a survey, or field research, for example, the techniques for subject selection will vary quite a bit. Happily, Chapter 7 of this book discusses sampling techniques for both qualitative and quantitative studies.

Measurement

What are the key variables in your study? How will you define and measure them? Do your definitions and measurement methods duplicate or differ from

those of previous research on this topic? If you have already developed your measurement device (a questionnaire, for example) or will be using something previously developed by others, it might be appropriate to include a copy in an appendix to your proposal.

Data-Collection Methods

How will you actually collect the data for your study? Will you conduct an experiment or a survey? Will you undertake field research or will you focus on the reanalysis of statistics already created by others? Perhaps you'll use more than one method.

Analysis

Indicate the kind of analysis you plan to conduct. Spell out the purpose and logic of your analysis. Are you interested in precise description? Do you intend to explain why things are the way they are? Do you plan to account for variations in some quality: for example, why some students are more liberal than others? What possible explanatory variables will your analysis consider, and how will you know if you've explained variations adequately?

Schedule

It's often appropriate to provide a schedule for the various stages of research. Even if you don't do this for the proposal, do it for yourself. Unless you have a timeline for accomplishing the several stages of research and keeping track of how you're doing, you may end up in trouble.

Budget

When you ask someone to cover the costs of your research, you need to provide a budget that specifies where the money will go. Large, expensive projects include budgetary categories such as personnel, equipment, supplies, telephones, and postage. Even for a project you'll pay for yourself, it's a good idea to spend some time anticipating expenses: office supplies, photocopying, CD-ROMs, telephone calls, transportation, and so on.

As you can see, if you're interested in conducting a social research project, it's a good idea to prepare a research proposal for your own purposes, even if you aren't required to do so by your instructor or a funding agency. If you're going to invest your time and energy in such a project, you should do what you can to insure a return on that investment.

Now that you've had a broad overview of social research, let's move on to the remaining chapters in this book and learn exactly how to design and execute each specific step. If you've found a research topic that really interests you, you'll want to keep it in mind as you see how you might go about studying it.

MAIN POINTS

Introduction

- Any research design requires researchers to specify as clearly as possible what they want to find out and then determine the best way to do it.

Three Purposes of Research

- The principal purposes of social research include exploration, description, and explanation. Research studies often combine more than one purpose.
- Exploration is the attempt to develop an initial, rough understanding of some phenomenon.
- Description is the precise measurement and reporting of the characteristics of some population or phenomenon under study.
- Explanation is the discovery and reporting of relationships among different aspects of the phenomenon under study. Whereas descriptive studies answer the question "What's so?" explanatory ones tend to answer the question "Why?"

The Logic of Nomothetic Explanation

- Both idiographic and nomothetic models of explanation rest on the idea of causation.

The idiographic model aims at a complete understanding of a particular phenomenon, using all relevant causal factors. The nomothetic model aims at a general understanding—not necessarily complete—of a class of phenomena, using a small number of relevant causal factors.

- There are three basic criteria for establishing causation in nomothetic analyses: (1) The variables must be empirically associated, or correlated, (2) the causal variable must occur earlier in time than the variable it is said to affect, and (3) the observed effect cannot be explained as the effect of a different variable.

Necessary and Sufficient Causes

- Mere association, or correlation, does not in itself establish causation. A spurious causal relationship is an association that in reality is caused by one or more other variables.

Units of Analysis

- Units of analysis are the people or things whose characteristics social researchers observe, describe, and explain. Typically, the unit of analysis in social research is the individual person, but it may also be a social group, a formal organization, a social interaction, a social artifact, or some other phenomenon such as a lifestyle or a type of social interaction.
- The ecological fallacy involves conclusions drawn from the analysis of groups (e.g., corporations) that are then assumed to apply to individuals (e.g., the employees of corporations).
- Reductionism is the attempt to understand a complex phenomenon in terms of a narrow set of concepts, such as attempting to explain the American Revolution solely in terms of economics (or political idealism or psychology).

The Time Dimension

- Research into processes that occur over time presents social challenges that can be addressed through cross-sectional studies or longitudinal studies.
- Cross-sectional studies are based on observations made at one time. Although such studies are limited by this characteristic, researchers can sometimes make inferences about processes that occur over time.
- In longitudinal studies, observations are made at many times. Such observations may be made of samples drawn from general populations (trend studies), samples drawn from more specific subpopulations (cohort studies), or the same sample of people each time (panel studies).

How to Design a Research Project

- Research design starts with an initial interest, idea, or theoretical expectation and proceeds through a series of interrelated steps to narrow the focus of the study so that concepts, methods, and procedures are well defined. A good research plan accounts for all these steps in advance.
- At the outset, a researcher specifies the meaning of the concepts or variables to be studied (conceptualization), chooses a research method or methods (e.g., experiments versus surveys), and specifies the population to be studied and, if applicable, how it will be sampled.
- To operationalize the concepts to be studied, the researcher states precisely how variables in the study will be measured. Research then proceeds through observation, data processing, analysis, and application, such as reporting the results and assessing their implications.

The Research Proposal

- A research proposal provides a preview of why a study will be undertaken and how it will be conducted. A research project is often required to get permission or necessary resources. Even when not required, a proposal is a useful device for planning.

KEY TERMS

The following terms are defined in context in the chapter and at the bottom of the page where the term

is introduced, as well as in the comprehensive glossary at the back of the book.

cohort study
correlation
cross-sectional study
ecological fallacy
longitudinal study
panel study
reductionism
social artifact
sociobiology
spurious relationship
trend study
units of analysis

REVIEW QUESTIONS AND EXERCISES

1. One example in this chapter suggested that political orientations cause attitudes toward legalizing marijuana. Can you make an argument that the time order is just the opposite of what was assumed?
2. Here are some examples of real research topics. For each one, can you name the unit of analysis? (The answers are at the end of this chapter.)
   a. Women watch TV more than men because they are likely to work fewer hours outside the home than men. . . . Black people watch an average of approximately three-quarters of an hour more television per day than white people. (Hughes 1980: 290)
   b. Of the 130 incorporated U.S. cities with more than 100,000 inhabitants in 1960, 126 had at least two short-term nonproprietary general hospitals accredited by the American Hospital Association. (Turk 1980: 317)
   c. The early TM [transcendental meditation] organizations were small and informal. The Los Angeles group, begun in June 1959, met at a member's house where, incidentally, Maharishi was living. (Johnston 1980: 337)
   d. However, it appears that the nursing staffs exercise strong influence over . . . a decision to change the nursing care system. . . . Conversely, among those decisions dominated by the administration and the medical staffs . . . (Comstock 1980: 77)
   e. Though 667,000 out of 2 million farmers in the United States are women, women historically have not been viewed as farmers, but rather, as the farmer's wife. (Votaw 1979: 8)
   f. The analysis of community opposition to group homes for the mentally handicapped . . . indicates that deteriorating neighborhoods are most likely to organize in opposition, but that upper-middle class neighborhoods are most likely to enjoy private access to local officials. (Graham and Hogan 1990: 513)
   g. Some analysts during the 1960s predicted that the rise of economic ambition and political militancy among blacks would foster discontent with the "otherworldly" black mainline churches. (Ellison and Sherkat 1990: 551)
   h. This analysis explores whether propositions and empirical findings of contemporary theories of organizations directly apply to both private product producing organizations (PPOs) and public human service organizations (PSOs). (Schiflett and Zey 1990: 569)
   i. This paper examines variations in job title structures across work roles. Analyzing 3,173 job titles in the California civil service system in 1985, we investigate how and why lines of work vary in the proliferation of job categories that differentiate ranks, functions, or particular organizational locations. (Strang and Baron 1990: 479)
3. Review the logic of spuriousness. Can you think up an example where an observed relationship between two variables could actually be explained away by a third variable?
4. Using InfoTrac College Edition or printed journals in the library, locate a research project involving a panel study. Describe the nature of the study design and its primary findings.

ADDITIONAL READINGS

Bart, Pauline, and Linda Frankel. 1986. *The Student Sociologist's Handbook*. Morristown, NJ: General Learning Press. A handy little reference book to help you get started on a research project. Written from the standpoint of a student term paper, this volume offers a particularly good guide to the periodical literature of the social sciences that's available in a good library.

Casley, D. J., and D. A. Lury. 1987. *Data Collection in Developing Countries*. Oxford: Clarendon Press.

This book discusses the special problems of research in the developing world.

Cooper, Harris M. 1989. *Integrating Research: A Guide for Literature Reviews*. Newbury Park, CA: Sage. The author leads you through each step in the literature review process.

Hunt, Morton. 1985. *Profiles of Social Research: The Scientific Study of Human Interactions*. New York: Basic Books. An engaging and informative series of project biographies: James Coleman's study of segregated schools is presented, as well as several other major projects that illustrate the elements of social research in actual practice.

Iversen, Gudmund R. 1991. *Contextual Analysis*. Newbury Park, CA: Sage. Contextual analysis examines the impact of socioenvironmental factors on individual behavior. Durkheim's study of suicide offers a good example of this, identifying social contexts that affect the likelihood of self-destruction.

Maxwell, Joseph A. 1996. *Qualitative Research Design: An Interactive Approach*. Newbury Park, CA: Sage. Maxwell covers many of the same topics that this chapter does but with attention devoted specifically to qualitative research projects.

Menard, Scott. 1991. *Longitudinal Research*. Newbury Park, CA: Sage. Beginning by explaining why researchers conduct longitudinal research, the author goes on to detail a variety of study designs as well as suggestions for the analysis of longitudinal data.

Miller, Delbert. 1991. *Handbook of Research Design and Social Measurement*. Newbury Park, CA: Sage. A useful reference book for introducing or reviewing numerous issues involved in design and measurement. In addition, the book contains a wealth of practical information relating to foundations, journals, and professional associations.

SPSS EXERCISES

See the booklet that accompanies your text for exercises using SPSS (Statistical Package for the Social Sciences). There are exercises offered for each chapter, and you'll also find a detailed primer on using SPSS.

Online Study Resources

SociologyNow™: Research Methods

1. Before you do your final review of the chapter, take the *SociologyNow: Research Methods* diagnostic quiz to help identify the areas on which you should concentrate. You'll find information on this online tool, as well as instructions on how to access all of its great resources, in the front of the book.
2. As you review, take advantage of the *Sociology Now: Research Methods* customized study plan, based on your quiz results. Use this study plan with its interactive exercises and other resources to master the material.
3. When you're finished with your review, take the posttest to confirm that you're ready to move on to the next chapter.

WEBSITE FOR *THE PRACTICE OF SOCIAL RESEARCH* 11TH EDITION

Go to your book's website at http://sociology.wadsworth.com/babbie_practice11e for tools to aid you in studying for your exams. You'll find *Tutorial Quizzes* with feedback, *Internet Exercises*, *Flashcards*, and *Chapter Tutorials*, as well as *Extended Projects*, *InfoTrac College Edition* search terms, *Social Research in Cyberspace*, *GSS Data*, *Web Links*, and primers for using various data-analysis software such as SPSS and NVivo.

WEB LINKS FOR THIS CHAPTER

Please realize that the Internet is an evolving entity, subject to change. Nevertheless, these few websites should be fairly stable. Also, check your book's website for even more *Web Links*.

The Internet Public Library, Social Sciences Resources
http://www.ipl.org/ref/RR/static/soc00.00.00.html
This site, along with its numerous hotlinks, provides a broad view of the kinds of research topics explored by social researchers in many disciplines.

University of Calgary, Beginner's Guide to the Research Proposal
http://www.ucalgary.ca/md/CAH/research/res_prop.htm

As the name suggests, this site will walk you through the process of preparing a research proposal.

Anthony W. Heath, "The Proposal in Qualitative Research"
http://www.nova.edu/ssss/QR/QR3-1/heath.html
This piece, reprinted from *The Qualitative Report* 3 (no. 1, March 1997) provides another guide to proposal writing, this time specifically for qualitative research projects.

ANSWERS TO UNITS OF ANALYSIS QUIZ, EXERCISE 2

a. Men and women, black and white people (individuals)
b. Incorporated U.S. cities (groups)
c. Transcendental meditation organizations (groups)
d. Nursing staffs (groups)
e. Farmers (individuals)
f. Neighborhoods (groups)
g. Blacks (individuals)
h. Service and production organizations (formal organizations)
i. Job titles (artifacts)

CHAPTER FIVE

Conceptualization, Operationalization, and Measurement

CHAPTER OVERVIEW

The interrelated steps of conceptualization, operationalization, and measurement allow researchers to turn a general idea for a research topic into useful and valid measurements in the real world. An essential part of this process involves transforming the relatively vague terms of ordinary language into precise objects of study with well-defined and measurable meanings.

SociologyNow™: Research Methods

Use this online tool to help you make the grade on your next exam. After reading this chapter, go to the "Online Study Resources" at the end of the chapter for instructions on how to benefit from *SociologyNow: Research Methods*.

Introduction

This chapter and the next deal with how researchers move from a general idea about what they want to study to effective and well-defined measurements in the real world. This chapter discusses the interrelated processes of conceptualization, operationalization, and measurement. Chapter 6 builds on this foundation to discuss types of measurements that are more complex.

We begin this chapter by confronting the hidden concern people sometimes have about whether it's truly possible to measure the stuff of life: love, hate, prejudice, religiosity, radicalism, alienation. The answer is yes, but it will take a few pages to see how. Once we establish that researchers can measure anything that exists, we'll turn to the steps involved in doing just that.

Measuring Anything That Exists

Earlier in this book, I said that one of the two pillars of science is observation. Because this word can suggest a casual, passive activity, scientists often use the term *measurement* instead, meaning careful, deliberate observations of the real world for the purpose of describing objects and events in terms of the attributes composing a variable.

You may have some reservations about the ability of science to measure the really important aspects of human social existence. If you've read research reports dealing with something like liberalism or religion or prejudice, you may have been dissatisfied with the way the researchers measured whatever they were studying. You may have felt that they were too superficial, that they missed the aspects that really matter most. Maybe they measured religiosity as the number of times a person went to religious services, or maybe they measured liberalism by how people voted in a single election. Your dissatisfaction would surely have increased if you had found yourself being misclassified by the measurement system.

Your feeling of dissatisfaction reflects an important fact about social research: Most of the variables we want to study don't actually exist in the way that rocks exist. Indeed, they are made up. Moreover, they seldom have a single, unambiguous meaning.

To see what I mean, suppose we want to study *political party affiliation*. To measure this variable, we might consult the list of registered voters to note whether the people we were studying were registered as Democrats or Republicans and take that as a measure of their party affiliation. But we could also simply ask someone what party they identify with and take their response as our measure. Notice that these two different measurement possibilities reflect somewhat different definitions of "political party affiliation." They might even produce different results: Someone may have registered as a Democrat years ago but gravitated more and more toward a Republican philosophy over time. Or someone who is registered with neither political party may, when asked, say she is affiliated with the one she feels the most kinship with.

Similar points apply to *religious affiliation*. Sometimes this variable refers to official membership in a particular church, temple, mosque, and so forth; other times it simply means whatever religion, if any, you identify yourself with. Perhaps to you it means something else, such as attendance at religious services.

The truth is that neither "party affiliation" nor "religious affiliation" has any *real* meaning, if by "real" we mean corresponding to some objective aspect of reality. These variables do not exist in nature. They are merely terms we've made up and assigned specific meanings to for some purpose, such as doing social research.

But, you might object, "political affiliation" and "religious affiliation"—and a host of other things social researchers are interested in, such as prejudice or compassion—have *some* reality. After all, we make statements about them, such as "In

Happytown, 55 percent of the adults affiliate with the Republican Party, and 45 percent of them are Episcopalians. Overall, people in Happytown are low in prejudice and high in compassion." Even ordinary people, not just social researchers, have been known to make statements like that. If these things don't exist in reality, what is it that we're measuring and talking about?

What indeed? Let's take a closer look by considering a variable of interest to many social researchers (and many other people as well)—*prejudice*.

Conceptions, Concepts, and Reality

As you and I wandered down the road of life, we observed a lot of things and knew they were real through our observations, and we heard reports from other people that seemed real. For example:

- We personally heard people say nasty things about minority groups.
- We heard people say that women were inferior to men.
- We read about African Americans being lynched.
- We read that women and minorities earned less for the same work.
- We learned about "ethnic cleansing" and wars in which one ethnic group tried to eradicate another.

With additional experience, we noticed something more. People who participated in lynching were also quite likely to call African Americans ugly names. A lot of them, moreover, seemed to want women to "stay in their place." Eventually it dawned on us that these several tendencies often appeared together in the same people and also had something in common. At some point, someone had a bright idea: "Let's use the word *prejudiced* as a shorthand notation for people like that. We can use the term even if they don't do all those things—as long as they're pretty much like that."

Being basically agreeable and interested in efficiency, we agreed to go along with the system. That's where "prejudice" came from. We never observed it. We just agreed to use it as a shortcut, a name that represents a collection of apparently related phenomena that we've each observed in the course of life. In short, we made it up.

Here's another clue that prejudice isn't something that exists apart from our rough agreement to use the term in a certain way. Each of us develops our own mental image of what the set of real phenomena we've observed represents in general and what these phenomena have in common. When I say the word *prejudice*, it evokes a mental image in your mind, just as it evokes one in mine. It's as though file drawers in our minds contained thousands of sheets of paper, with each sheet of paper labeled in the upper right-hand corner. A sheet of paper in each of our minds has the term *prejudice* on it. On your sheet are all the things you've been told about prejudice and everything you've observed that seems to be an example of it. My sheet has what I've been told about it plus all the things I've observed that seem examples of it—and mine isn't the same as yours.

The technical term for those mental images, those sheets of paper in our mental file drawers, is *conception*. That is, I have a conception of prejudice, and so do you. We can't communicate these mental images directly, so we use the terms written in the upper right-hand corner of our own mental sheets of paper as a way of communicating about our conceptions and the things we observe that are related to those conceptions. These terms make it possible for us to communicate and eventually agree on what we specifically mean by those terms. In social research, the process of coming to an agreement about what terms mean is *conceptualization*, and the result is called a *concept*.

Let's take another example of a conception. Suppose that I'm going to meet someone named Pat, whom you already know. I ask you what Pat is like. Now suppose that you've seen Pat help lost children find their parents and put a tiny bird back in its nest. Pat got you to take turkeys to poor families on Thanksgiving and to visit a children's hospital on Christmas. You've seen Pat weep through a movie about a mother overcoming adversities to save and protect her child. As you search through your mental files, you may find all or most of those

phenomena recorded on a single sheet labeled "compassionate." You look over the other entries on the page, and you find they seem to provide an accurate description of Pat. So you say, "Pat is compassionate."

Now I leaf through my own mental file drawer until I find a sheet marked "compassionate." I then look over the things written on my sheet, and I say, "Oh, that's nice." I now feel I know what Pat is like, but my expectations reflect the entries on my file sheet, not yours. Later, when I meet Pat, I happen to find that my own experiences correspond to the entries I have on my "compassionate" file sheet, and I say that you sure were right.

But suppose my observations of Pat contradict the things I have on my file sheet. I tell you that I don't think Pat is very compassionate, and we begin to compare notes.

You say, "I once saw Pat weep through a movie about a mother overcoming adversity to save and protect her child." I look at my "compassionate sheet" and can't find anything like that. Looking elsewhere in my file, I locate that sort of phenomenon on a sheet labeled "sentimental." I retort, "That's not compassion. That's just sentimentality."

To further strengthen my case, I tell you that I saw Pat refuse to give money to an organization dedicated to saving whales from extinction. "That represents a lack of compassion," I argue. You search through your files and find saving the whales on two sheets—"environmental activism" and "cross-species dating"—and you say so. Eventually, we set about comparing the entries we have on our respective sheets labeled "compassionate." We then discover that many of our mental images corresponding to that term differ.

In the big picture, language and communication work only to the extent that you and I have considerable overlap in the kinds of entries we have on our corresponding mental file sheets. The similarities we have on those sheets represent the agreements existing in our society. As we grow up, we're told approximately the same thing when we're first introduced to a particular term. Dictionaries formalize the agreements our society has about such terms. Each of us, then, shapes his or her mental images to correspond with such agreements. But because all of us have different experiences and observations, no two people end up with exactly the same set of entries on any sheet in their file systems. If we want to measure "prejudice" or "compassion," we must first stipulate what, exactly, counts as prejudice or compassion for our purposes.

Returning to the assertion made at the outset of this chapter, we can measure anything that's real. We can measure, for example, whether Pat actually puts the little bird back in its nest, visits the hospital on Christmas, weeps at the movie, or refuses to contribute to saving the whales. All of those behaviors exist, so we can measure them. But is Pat really compassionate? We can't answer that question; we can't measure compassion in any objective sense, because compassion doesn't exist in the way that those things I just described exist. Compassion exists only in the form of the agreements we have about how to use the term in communicating about things that are real.

Concepts as Constructs

If you recall the discussions of postmodernism in Chapter 1, you'll recognize that some people would object to the degree of "reality" I've allowed in the preceding comments. Did Pat "really" visit the hospital on Christmas? Does the hospital "really" exist? Does Christmas? Though we aren't going to be radically postmodern in this chapter, I think you'll recognize the importance of an intellectually tough view of what's real and what's not. (When the intellectual going gets tough, the tough become social scientists.)

In this context, Abraham Kaplan (1964) distinguishes three classes of things that scientists measure. The first class is *direct observables*: those things we can observe rather simply and directly, like the color of an apple or the check mark made in a questionnaire. The second class, *indirect observables*, require "relatively more subtle, complex, or indirect observations" (1964: 55). We note a person's check mark beside "female" in a questionnaire and have indirectly observed that person's gender. History books or minutes of corporate board meetings provide indirect observations of past social actions. Finally, the third class of observables consists of

constructs—theoretical creations that are based on observations but that cannot be observed directly or indirectly. A good example is intelligence quotient, or IQ. It is *constructed* mathematically from observations of the answers given to a large number of questions on an IQ test. No one can directly or indirectly observe IQ. It is no more a "real" characteristic of people than is compassion or prejudice.

Kaplan (1964: 49) defines *concept* as a "family of conceptions." A concept is, as Kaplan notes, a construct, something we create. Concepts such as compassion and prejudice are constructs created from your conception of them, my conception of them, and the conceptions of all those who have ever used these terms. They cannot be observed directly or indirectly, because they don't exist. We made them up.

To summarize, concepts are constructs derived by mutual agreement from mental images (conceptions). Our conceptions summarize collections of seemingly related observations and experiences. Although the observations and experiences are real, at least subjectively, conceptions, and the concepts derived from them, are only mental creations. The terms associated with concepts are merely devices created for the purposes of filing and communication. A term such as *prejudice* is, objectively speaking, only a collection of letters. It has no intrinsic reality beyond that. Is has only the meaning we agree to give it.

Usually, however, we fall into the trap of believing that terms for constructs do have intrinsic meaning, that they name real entities in the world. That danger seems to grow stronger when we begin to take terms seriously and attempt to use them precisely. Further, the danger is all the greater in the presence of experts who appear to know more than we do about what the terms really mean: It's easy to yield to authority in such a situation.

Once we assume that terms like *prejudice* and *compassion* have real meanings, we begin the tortured task of discovering what those real meanings are and what constitutes a genuine measurement of them. Regarding constructs as real is called *reification*. The reification of concepts in day-to-day life is quite common. In science, we want to be quite clear about what it is we are actually measuring, but this aim brings a pitfall with it. Settling on the "best" way of measuring a variable in a particular study may imply that we've discovered the "real" meaning of the concept involved. In fact, concepts have no real, true, or objective meanings—only those we agree are best for a particular purpose.

Does this discussion imply that compassion, prejudice, and similar constructs can't be measured? Interestingly, the answer is no. (And a good thing, too, or a lot of us social researcher types would be out of work.) I've said that we can measure anything that's real. Constructs aren't real in the way that trees are real, but they do have another important virtue: They are useful. That is, they help us organize, communicate about, and understand things that *are* real. They help us make predictions about real things. Some of those predictions even turn out to be true. Constructs can work this way because, although not real or observable in themselves, they have a definite relationship to things that *are* real and observable. The bridge from direct and indirect observables to useful constructs is the process called conceptualization.

conceptualization The mental process whereby fuzzy and imprecise notions (concepts) are made more specific and precise. So you want to study prejudice. What do you mean by "prejudice"? Are there different kinds of prejudice? What are they?

Conceptualization

As we've seen, day-to-day communication usually occurs through a system of vague and general agreements about the use of terms. Although you and I do not agree completely about the use of the term *compassionate*, I'm probably safe in assuming that Pat won't pull the wings off flies. A wide range of misunderstandings and conflict—from the interpersonal to the international—is the price we pay for our imprecision, but somehow we muddle through. Science, however, aims at more than muddling; it cannot operate in a context of such imprecision.

The process through which we specify what we mean when we use particular terms in research is called **conceptualization**. Suppose we want to

find out, for example, whether women are more compassionate than men. I suspect many people assume this is the case, but it might be interesting to find out if it's really so. We can't meaningfully study the question, let alone agree on the answer, without some working agreements about the meaning of *compassion*. They are "working" agreements in the sense that they allow us to work on the question. We don't need to agree or even pretend to agree that a particular specification is ultimately the best one.

Conceptualization, then, produces a specific, agreed-on meaning for a concept for the purposes of research. This process of specifying exact meaning involves describing the indicators we'll be using to measure our concept and the different aspects of the concept, called dimensions.

Indicators and Dimensions

Conceptualization gives definite meaning to a concept by specifying one or more indicators of what we have in mind. An **indicator** is a sign of the presence or absence of the concept we're studying. Here's an example.

We might agree that visiting children's hospitals during Christmas and Hanukkah is an indicator of compassion. Putting little birds back in their nests might be agreed on as another indicator, and so forth. If the unit of analysis for our study is the individual person, we can then observe the presence or absence of each indicator for each person under study. Going beyond that, we can add up the number of indicators of compassion observed for each individual. We might agree on ten specific indicators, for example, and find six present in our study of Pat, three for John, nine for Mary, and so forth.

Returning to our question about whether men or women are more compassionate, we might calculate that the women we studied displayed an average of 6.5 indicators of compassion, the men an average of 3.2. On the basis of our quantitative analysis of group difference, we might therefore conclude that women are, on the whole, more compassionate than men.

Usually, though, it's not that simple. Imagine you're interested in understanding a small fundamentalist religious cult, particularly their harsh views on various groups: gays, nonbelievers, feminists, and others. In fact, they suggest that anyone who refuses to join their group and abide by its teachings will "burn in hell." In the context of your interest in compassion, they don't seem to have much. And yet, the group's literature often speaks of their compassion for others. You want to explore this seeming paradox.

To pursue this research interest, you might arrange to interact with cult members, getting to know them and learning more about their views. You could tell them you were a social researcher interested in learning about their group, or perhaps you would just express an interest in learning more, without saying why.

In the course of your conversations with group members and perhaps attendance of religious services, you would put yourself in situations where you could come to understand what the cult members mean by *compassion*. You might learn, for example, that members of the group were so deeply concerned about sinners burning in hell that they were willing to be aggressive, even violent, to make people change their sinful ways. Within their own paradigm, then, cult members would see beating up gays, prostitutes, and abortion doctors as acts of compassion.

Social researchers focus their attention on the meanings that the people under study give to words and actions. Doing so can often clarify the behaviors observed: At least now you understand how the cult can see violent acts as compassionate. On the other hand, paying attention to what words and actions mean to the people under study almost always complicates the concepts researchers are interested in. (We'll return to this issue when we discuss the validity of measures, toward the end of this chapter.)

Whenever we take our concepts seriously and set about specifying what we mean by them, we

indicator An observation that we choose to consider as a reflection of a variable we wish to study. Thus, for example, attending religious services might be considered an indicator of *religiosity*.

discover disagreements and inconsistencies. Not only do you and I disagree, but each of us is likely to find a good deal of muddiness within our own mental images. If you take a moment to look at what you mean by compassion, you'll probably find that your image contains several kinds of compassion. That is, the entries on your mental file sheet can be combined into groups and subgroups, say, compassion toward friends, co-religionists, humans, and birds. You may also find several different strategies for making combinations. For example, you might group the entries into feelings and actions.

The technical term for such groupings is **dimension**: a specifiable aspect of a concept. For instance, we might speak of the "feeling dimension" of compassion and the "action dimension" of compassion. In a different grouping scheme, we might distinguish "compassion for humans" from "compassion for animals." Or we might see compassion as helping people have what we want for them versus what they want for themselves. Still differently, we might distinguish compassion as forgiveness from compassion as pity.

Thus, we could subdivide compassion into several clearly defined dimensions. A complete conceptualization involves both specifying dimensions and identifying the various indicators for each.

Sometimes conceptualization aimed at identifying different dimensions of a variable leads to a different kind of distinction. We may conclude that we've been using the same word for meaningfully distinguishable concepts. In the following example, the researchers find (1) that "violence" is not a sufficient description of "genocide" and (2) that the concept "genocide" itself comprises several distinct phenomena. Let's look at the process they went through to come to this conclusion.

When Daniel Chirot and Jennifer Edwards attempted to define the concept of "genocide," they found existing assumptions were not precise enough for their purposes:

> The United Nations originally defined it as an attempt to destroy "in whole or in part, a national, ethnic, racial, or religious group." If genocide is distinct from other types of violence, it requires its own unique explanation.
>
> *(2003: 14)*

Notice the final comment in this excerpt, as it provides an important insight into why researchers are so careful in specifying the concepts they study. If genocide, such as the Holocaust, were simply another example of violence, like assaults and homicides, then what we know about violence in general might explain genocide. If it differs from other forms of violence, then we may need a different explanation for it. So, the researchers began by suggesting that "genocide" was a concept distinct from "violence" for their purposes.

Then, as Chirot and Edwards examined historical instances of genocide, they began concluding that the motivations for launching genocidal mayhem differed sufficiently to represent four distinct phenomena that were all called "genocide" (2003: 15–18).

1. *Convenience*: Sometimes the attempt to eradicate a group of people serves a function for the eradicators, such as Julius Caesar's attempt to eradicate tribes defeated in battle, fearing they would be difficult to rule. Or when gold was discovered on Cherokee land in the Southeastern United States in the early nineteenth century, the Cherokee were forcibly relocated to Oklahoma in an event known as the "Trail of Tears," which ultimately killed as many as half of those forced to leave.
2. *Revenge*: When the Chinese of Nanking bravely resisted the Japanese invaders in the early years of World War II, the conquerors felt they had been insulted by those they regarded as inferior beings. Tens of thousands were slaughtered in the "Rape of Nanking" in 1937–1938.
3. *Fear*: The ethnic cleansing that recently occurred in the former Yugoslavia was at least

dimension A specifiable aspect of a concept. "Religiosity," for example, might be specified in terms of a belief dimension, a ritual dimension, a devotional dimension, a knowledge dimension, and so forth.

partly motivated by economic competition and worries that the growing Albanian population of Kosovo was gaining political strength through numbers. Similarly, the Hutu attempt to eradicate the Tutsis of Rwanda grew out of a fear that returning Tutsi refugees would seize control of the country. Often intergroup fears such as these grow out of long histories of atrocities, often inflicted in both directions.

4. *Purification*: The Nazi Holocaust, probably the most publicized case of genocide, was intended as a purification of the "Aryan race." While Jews were the main target, gypsies, homosexuals, and many other groups were also included. Other examples include the Indonesian witch-hunt against communists in 1965–1966 and the attempt to eradicate all non-Khmer Cambodians under Pol Pot in the 1970s.

No single theory of genocide could explain these various forms of mayhem. Indeed, this act of conceptualization suggests four distinct phenomena, each needing a different set of explanations.

Specifying the different dimensions of a concept often paves the way for a more sophisticated understanding of what we're studying. We might observe, for example, that women are more compassionate in terms of feelings, and men more so in terms of actions—or vice versa. Whichever turned out to be the case, we would not be able to say whether men or women are really more compassionate. Our research would have shown that there is no single answer to the question. That alone represents an advance in our understanding of reality. To get a better feel for concepts, variables, and indicators, go to the General Social Survey codebook and explore some of the ways the researchers have measured various concepts: http:// www.icpsr .umich.edu/GSS99/subject/s-index.htm.

The Interchangeability of Indicators

There is another way that the notion of indicators can help us in our attempts to understand reality by means of "unreal" constructs. Suppose, for the moment, that you and I have compiled a list of 100 indicators of compassion and its various dimensions. Suppose further that we disagree widely on which indicators give the clearest evidence of compassion or its absence. If we pretty much agree on some indicators, we could focus our attention on those, and we would probably agree on the answer they provided. We would then be able to say that some people are more compassionate than others in some dimension. But suppose we don't really agree on any of the possible indicators. Surprisingly, we can still reach an agreement on whether men or women are the more compassionate. How we do that has to do with the interchangeability of indicators.

The logic works like this. If we disagree totally on the value of the indicators, one solution would be to study all of them. Suppose that women turn out to be more compassionate than men on all 100 indicators—on all the indicators you favor and on all of mine. Then we would be able to agree that women are more compassionate than men, even though we still disagree on exactly what compassion means in general.

The interchangeability of indicators means that if several different indicators all represent, to some degree, the same concept, then all of them will behave the same way that the concept would behave if it were real and could be observed. Thus, given a basic agreement about what "compassion" is, if women are generally more compassionate than men, we should be able to observe that difference by using any reasonable measure of compassion. If, on the other hand, women are more compassionate than men on some indicators but not on others, we should see if the two sets of indicators represent different dimensions of compassion.

You have now seen the fundamental logic of conceptualization and measurement. The discussions that follow are mainly refinements and extensions of what you've just read. Before turning to a technical elaboration of measurement, however, we need to fill out the picture of conceptualization by looking at some of the ways social researchers provide standards, consistency, and commonality for the meanings of terms.

Real, Nominal, and Operational Definitions

As we have seen, the design and execution of social research requires us to clear away the confusion over concepts and reality. To this end, logicians and scientists have found it useful to distinguish three kinds of definitions: real, nominal, and operational.

The first of these reflects the reification of terms. As Carl Hempel cautions,

> A "real" definition, according to traditional logic, is not a stipulation determining the meaning of some expression but a statement of the "essential nature" or the "essential attributes" of some entity. The notion of essential nature, however, is so vague as to render this characterization useless for the purposes of rigorous inquiry.
>
> *(1952: 6)*

In other words, trying to specify the "real" meaning of concepts only leads to a quagmire: It mistakes a construct for a real entity.

The **specification** of concepts in scientific inquiry depends instead on nominal and operational definitions. A *nominal definition* is one that is simply assigned to a term without any claim that the definition represents a "real" entity. Nominal definitions are arbitrary—I could define *compassion* as "plucking feathers off helpless birds" if I wanted to—but they can be more or less useful. For most purposes, especially communication, that last definition of compassion would be pretty useless. Most nominal definitions represent some consensus, or convention, about how a particular term is to be used.

An *operational definition*, as you may remember from an earlier chapter, specifies precisely how a concept will be measured—that is, the operations we'll perform. An operational definition is nominal rather than real, but it has the advantage of achieving maximum clarity about what a concept means in the context of a given study. In the midst of disagreement and confusion over what a term "really" means, we can specify a working definition for the purposes of an inquiry. Wishing to examine socioeconomic status (SES) in a study, for example, we may simply specify that we are going to treat SES as a combination of income and educational attainment. In this decision, we rule out other possible aspects of SES: occupational status, money in the bank, property, lineage, lifestyle, and so forth. Our findings will then be interesting to the extent that our definition of SES is useful for our purpose.

specification The process through which concepts are made more specific.

Creating Conceptual Order

The clarification of concepts is a continuing process in social research. Catherine Marshall and Gretchen Rossman (1995: 18) speak of a "conceptual funnel" through which a researcher's interest becomes increasingly focused. Thus, a general interest in social activism could narrow to "individuals who are committed to empowerment and social change" and further focus on discovering "what experiences shaped the development of fully committed social activists." This focusing process is inescapably linked to the language we use.

In some forms of qualitative research, the clarification of concepts is a key element in the collection of data. Suppose you were conducting interviews and observations in a radical political group devoted to combating oppression in U.S. society. Imagine how the meaning of oppression would shift as you delved more and more deeply into the members' experiences and worldviews. For example, you might start out thinking of oppression in physical and perhaps economic terms. The more you learned about the group, however, the more you might appreciate the possibility of psychological oppression.

The same point applies even to contexts where meanings might seem more fixed. In the analysis of textual materials, for example, social researchers sometimes speak of the "hermeneutic circle," a cyclical process of ever-deeper understanding.

> The understanding of a text takes place through a process in which the meaning of the separate parts is determined by the global meaning of

> the text as it is anticipated. The closer determination of the meaning of the separate parts may eventually change the originally anticipated meaning of the totality, which again influences the meaning of the separate parts, and so on.
>
> *(Kvale 1996: 47)*

Consider the concept "prejudice." Suppose you needed to write a definition of the term. You might start out thinking about racial/ethnic prejudice. At some point you would realize you should probably allow for gender prejudice, religious prejudice, antigay prejudice, and the like in your definition. Examining each of these specific types of prejudice would affect your overall understanding of the general concept. As your general understanding changed, however, you would likely see each of the individual forms somewhat differently.

The continual refinement of concepts occurs in all social research methods. Often you will find yourself refining the meaning of important concepts even as you write up your final report.

Although conceptualization is a continuing process, it is vital to address it specifically at the beginning of any study design, especially rigorously structured research designs such as surveys and experiments. In a survey, for example, operationalization results in a commitment to a specific set of questionnaire items that will represent the concepts under study. Without that commitment, the study could not proceed.

Even in less-structured research methods, however, it's important to begin with an initial set of anticipated meanings that can be refined during data collection and interpretation. No one seriously believes we can observe life with no preconceptions; for this reason, scientific observers must be conscious of and explicit about these conceptual starting points.

Let's explore initial conceptualization the way it applies to structured inquiries such as surveys and experiments. Though specifying nominal definitions focuses our observational strategy, it does not allow us to observe. As a next step we must specify exactly what we are going to observe, how we will do it, and what interpretations we are going to place on various possible observations. All these further specifications make up the operational definition of the concept.

In the example of socioeconomic status, we might decide to ask survey respondents two questions, corresponding to the decision to measure SES in terms of income and educational attainment:

1. What was your total family income during the past 12 months?
2. What is the highest level of school you completed?

To organize our data, we'd probably want to specify a system for categorizing the answers people give us. For income, we might use categories such as "under $5,000," "$5,000 to $10,000," and so on. Educational attainment might be similarly grouped in categories: less than high school, high school, college, graduate degree. Finally, we would specify the way a person's responses to these two questions would be combined in creating a measure of SES.

In this way we would create a working and workable definition of SES. Although others might disagree with our conceptualization and operationalization, the definition would have one essential scientific virtue: It would be absolutely specific and unambiguous. Even if someone disagreed with our definition, that person would have a good idea how to interpret our research results, because what we meant by SES—reflected in our analyses and conclusions—would be precise and clear.

Here is a diagram showing the progression of measurement steps from our vague sense of what a term means to specific measurements in a fully structured scientific study:

Conceptualization
↓
Nominal Definition
↓
Operational Definition
↓
Measurements in the Real World

An Example of Conceptualization: The Concept of Anomie

To bring this discussion of conceptualization in research together, let's look briefly at the history of a specific social scientific concept. Researchers studying urban riots are often interested in the part played by feelings of powerlessness. Social scientists sometimes use the word *anomie* in this context. This term was first introduced into social science by Emile Durkheim, the great French sociologist, in his classic 1897 study, *Suicide*.

Using only government publications on suicide rates in different regions and countries, Durkheim produced a work of analytic genius. To determine the effects of religion on suicide, he compared the suicide rates of predominantly Protestant countries with those of predominantly Catholic ones, Protestant regions of Catholic countries with Catholic regions of Protestant countries, and so forth. To determine the possible effects of the weather, he compared suicide rates in northern and southern countries and regions, and he examined the different suicide rates across the months and seasons of the year. Thus, he could draw conclusions about a supremely individualistic and personal act without having any data about the individuals engaging in it.

At a more general level, Durkheim suggested that suicide also reflects the extent to which a society's agreements are clear and stable. Noting that times of social upheaval and change often present individuals with grave uncertainties about what is expected of them, Durkheim suggested that such uncertainties cause confusion, anxiety, and even self-destruction. To describe this societal condition of normlessness, Durkheim chose the term *anomie*. Durkheim did not make this word up. Used in both German and French, it literally meant "without law." The English term *anomy* had been used for at least three centuries before Durkheim to mean disregard for divine law. However, Durkheim created the social scientific concept of anomie.

In the years that have followed the publication of *Suicide*, social scientists have found anomie a useful concept, and many have expanded on Durkheim's use. Robert Merton, in a classic article entitled "Social Structure and Anomie" (1938), concluded that anomie results from a disparity between the goals and means prescribed by a society. Monetary success, for example, is a widely shared goal in our society, yet not all individuals have the resources to achieve it through acceptable means. An emphasis on the goal itself, Merton suggested, produces normlessness, because those denied the traditional avenues to wealth go about getting it through illegitimate means. Merton's discussion, then, could be considered a further conceptualization of the concept of anomie.

Although Durkheim originally used the concept of anomie as a characteristic of societies, as did Merton after him, other social scientists have used it to describe individuals. To clarify this distinction, some scholars have chosen to use *anomie* in reference to its original, societal meaning and to use the term *anomia* in reference to the individual characteristic. In a given society, then, some individuals experience anomia, and others do not. Elwin Powell, writing 20 years after Merton, provided the following conceptualization of anomia (though using the term *anomie*) as a characteristic of individuals:

> When the ends of action become contradictory, inaccessible or insignificant, a condition of anomie arises. Characterized by a general loss of orientation and accompanied by feelings of "emptiness" and apathy, anomie can be simply conceived as meaninglessness.
>
> *(1958: 132)*

Powell went on to suggest there were two distinct kinds of anomia and to examine how the two rose out of different occupational experiences to result at times in suicide. In his study, however, Powell did not measure anomia per se; he studied the relationship between suicide and occupation, making inferences about the two kinds of anomia. Thus, the study did not provide an operational definition of *anomia*, only a further conceptualization.

Although many researchers have offered operational definitions of *anomia*, one name stands out

The Origins of Anomia

by Leo Srole

My career-long fixation on anomie began with reading Durkheim's *Le Suicide* as a Harvard undergraduate. Later, as a graduate student at Chicago, I studied under two Durkheimian anthropologists: William Lloyd Warner and Alfred Radcliffe-Brown. Radcliffe-Brown had carried on a lively correspondence with Durkheim, making me a collateral "descendant" of the great French sociologist.

For me, the early impact of Durkheim's work on suicide was mixed but permanent. On the one hand, I had serious reservations about his strenuous, ingenious, and often awkward efforts to force the crude, bureaucratic records on suicide rates to fit with his unidirectional sociological determinism. On the other hand, I was moved by Durkheim's unswerving preoccupation with the moral force of the interpersonal ties that bind us to our time, place, and past, and also his insights about the lethal consequences that can follow from shrinkage and decay in those ties.

My interest in anomie received an eyewitness jolt at the finale of World War II, when I served with the United Nations Relief and Rehabilitation Administration, helping to rebuild a war-torn Europe. At the Nazi concentration camp of Dachau, I saw firsthand the depths of dehumanization that macrosocial forces, such as those that engaged Durkheim, could produce in individuals like Hitler, Eichmann, and the others serving their dictates at all levels in the Nazi death factories.

Returning from my UNRRA post, I felt most urgently that the time was long overdue to come to an understanding of the dynamics underlying disintegrated social bonds. We needed to work expeditiously, deemphasizing proliferation of macro-level theory in favor of a direct exploratory encounter with individuals, using newly developed state-of-the-art survey research methodology. Such research, I also felt, should focus on a broader spectrum of behavioral pathologies than suicide.

My initial investigations were a diverse effort. In 1950, for example, I was able to interview a sample of 401 bus riders in Springfield, Massachusetts. Four years later, the Midtown Manhattan Mental Health Study provided a much larger population reach. These and other field projects gave me scope to expand and refine my measurements of that quality in individuals which reflected the macro-social quality Durkheim had called *anomie*.

While I began by using Durkheim's term in my own work, I soon decided that it was necessary to limit the use of that concept to its macro-social meaning and to sharply segregate it from its individual manifestations. For the latter purpose, the cognate but hitherto obsolete Greek term, *anomia*, readily suggested itself.

I first published the anomia construct in a 1956 article in the *American Sociological Review*, describing ways of operationalizing it, and presenting the results of its initial field application research. By 1982, the Science Citation Index and Social Science Citation Index had listed some 400 publications in political science, psychology, social work, and sociology journals here and abroad that had cited use of that article's instruments or findings, warranting the American Institute for Scientific Information to designate it a "citation classic."

over all. Two years before Powell's article appeared, Leo Srole (1956) published a set of questionnaire items that he said provided a good measure of anomia as experienced by individuals. It consists of five statements that subjects were asked to agree or disagree with:

1. In spite of what some people say, the lot of the average man is getting worse.
2. It's hardly fair to bring children into the world with the way things look for the future.
3. Nowadays a person has to live pretty much for today and let tomorrow take care of itself.
4. These days a person doesn't really know who he can count on.
5. There's little use writing to public officials because they aren't really interested in the problems of the average man.

(1956: 713)

In the half-century following its publication, the Srole scale has become a research staple for social scientists. You'll likely find this particular operationalization of anomia used in many of the research projects reported in academic journals. Srole touches on this in the accompanying box, "The Origins of Anomia," which he prepared for this book before his death.

This abbreviated history of anomie and anomia as social scientific concepts illustrates several points.

First, it's a good example of the process through which general concepts become operationalized measurements. This is not to say that the issue of how to operationalize anomie/anomia has been resolved once and for all. Scholars will surely continue to reconceptualize and reoperationalize these concepts for years to come, continually seeking more-useful measures.

The Srole scale illustrates another important point. Letting conceptualization and operationalization be open-ended does not necessarily produce anarchy and chaos, as you might expect. Order often emerges. For one thing, although we could define *anomia* any way we chose—in terms of, say, shoe size—we're likely to define it in ways not too different from other people's mental images. If you were to use a really offbeat definition, people would probably ignore you.

A second source of order is that, as researchers discover the utility of a particular conceptualization and operationalization of a concept, they're likely to adopt it, which leads to standardized definitions of concepts. Besides the Srole scale, examples include IQ tests and a host of demographic and economic measures developed by the U.S. Census Bureau. Using such established measures has two advantages: They have been extensively pretested and debugged, and studies using the same scales can be compared. If you and I do separate studies of two different groups and use the Srole scale, we can compare our two groups on the basis of anomia.

Social scientists, then, can measure anything that's real; through conceptualization and operationalization, they can even do a pretty good job of measuring things that aren't. Granting that such concepts as socioeconomic status, prejudice, compassion, and anomia aren't ultimately real, social scientists can create order in handling them. It is an order based on utility, however, not on ultimate truth.

Definitions in Descriptive and Explanatory Studies

As you'll recall from Chapter 4, two general purposes of research are description and explanation. The distinction between them has important implications for definition and measurement. If it seems that description is simpler than explanation, you may be surprised to learn that definitions are more problematic for descriptive research than for explanatory research. Before we turn to other aspects of measurement, you'll need a basic understanding of why this is so (we'll discuss this point more fully in Part 4).

It's easy to see the importance of clear and precise definitions for descriptive research. If we want to describe and report the unemployment rate in a city, our definition of being unemployed is obviously critical. That definition will depend on our definition of another term: the labor force. If it seems patently absurd to regard a three-year-old child as being unemployed, it is because such a child is not considered a member of the labor force. Thus, we might follow the U.S. Census Bureau's convention and exclude all people under 14 years of age from the labor force.

This convention alone, however, would not give us a satisfactory definition, because it would count as unemployed such people as high school students, the retired, the disabled, and homemakers. We might follow the census convention further by defining the labor force as "all persons 14 years of age and over who are employed, looking for work, or waiting to be called back to a job from which they have been laid off or furloughed." If a student, homemaker, or retired person is not looking for work, such a person would not be included in the labor force. Unemployed people, then, would be those members of the labor force, as defined, who are not employed.

But what does "looking for work" mean? Must a person register with the state employment service or go from door to door asking for employment? Or would it be sufficient to want a job or be open to an offer of employment? Conventionally, "looking for work" is defined operationally as saying yes in response to an interviewer's asking "Have you been looking for a job during the past seven days?" (Seven days is the period most often specified, but for some research purposes it might make more sense to shorten or lengthen it.)

As you can see, the conclusion of a descriptive study about the unemployment rate depends directly on how each issue of definition is resolved.

Increasing the period during which people are counted as looking for work would add more unemployed people to the labor force as defined, thereby increasing the reported unemployment rate. If we follow another convention and speak of the civilian labor force and the civilian unemployment rate, we're excluding military personnel; that, too, increases the reported unemployment rate, because military personnel would be employed—by definition. Thus, the descriptive statement that the unemployment rate in a city is 3 percent, or 9 percent, or whatever it might be, depends directly on the operational definitions used.

This example is relatively clear because there are several accepted conventions relating to the labor force and unemployment. Now, consider how difficult it would be to get agreement about the definitions you would need in order to say, "Forty-five percent of the students at this institution are politically conservative." Like the unemployment rate, this percentage would depend directly on the definition of what is being measured—in this case, political conservatism. A different definition might result in the conclusion "Five percent of the student body are politically conservative."

Ironically, definitions are less problematic in the case of explanatory research. Let's suppose we're interested in explaining political conservatism. Why are some people conservative and others not? More specifically, let's suppose we're interested in whether conservatism increases with age. What if you and I have 25 different operational definitions of *conservative*, and we can't agree on which definition is best? As we saw in the discussion of indicators, this is not necessarily an insurmountable obstacle to our research. Suppose we found old people to be more conservative than young people in terms of all 25 definitions. Clearly, the exact definition wouldn't matter much. We would conclude that old people are generally more conservative than young people—even though we couldn't agree about exactly what *conservative* means.

In practice, explanatory research seldom results in findings quite as unambiguous as this example suggests; nonetheless, the general pattern is quite common in actual research. There are consistent patterns of relationships in human social life that result in consistent research findings. However, such consistency does not appear in a descriptive situation. Changing definitions almost inevitably results in different descriptive conclusions. "The Importance of Variable Names" explores this issue in connection with the variable *citizen participation*.

Operationalization Choices

In discussing conceptualization, I frequently have referred to operationalization, for the two are intimately linked. To recap: Conceptualization is the refinement and specification of abstract concepts, and operationalization is the development of specific research procedures (operations) that will result in empirical observations representing those concepts in the real world.

As with the methods of data collection, social researchers have a variety of choices when operationalizing a concept. Although the several choices are intimately interconnected, I've separated them for the sake of discussion. Realize, though, that operationalization does not proceed through a systematic checklist.

Range of Variation

In operationalizing any concept, researchers must be clear about the range of variation that interests them. The question is, to what extent are they willing to combine attributes in fairly gross categories?

Let's suppose you want to measure people's incomes in a study by collecting the information from either records or interviews. The highest annual incomes people receive run into the millions of dollars, but not many people get that much. Unless you're studying the very rich, it probably won't add much to your study to keep track of extremely high categories. Depending on whom you study, you'll probably want to establish a highest income category with a much lower floor—maybe $100,000 or more. Although this decision will lead you to throw together people who earn a trillion dollars a year with paupers earning a mere $100,000, they'll survive it, and that mixing probably won't hurt your research any, either. The same decision faces you at

The Importance of Variable Names

by Patricia Fisher

Graduate School of Planning, University of Tennessee

Operationalization is one of those things that's easier said than done. It is quite simple to explain to someone the purpose and importance of operational definitions for variables, and even to describe how operationalization typically takes place. However, until you've tried to operationalize a rather complex variable, you may not appreciate some of the subtle difficulties involved. Of considerable importance to the operationalization effort is the particular name that you have chosen for a variable. Let's consider an example from the field of Urban Planning.

A variable of interest to planners is *citizen participation*. Planners are convinced that participation in the planning process by citizens is important to the success of plan implementation. Citizen participation is an aid to planners' understanding of the real and perceived needs of a community, and such involvement by citizens tends to enhance their cooperation with and support for planning efforts. Although many different conceptual definitions might be offered by different planners, there would be little misunderstanding over what is meant by *citizen participation*. The name of the variable seems adequate.

However, if we ask different planners to provide very simple operational measures for citizen participation, we are likely to find a variety among their responses that does generate confusion. One planner might keep a tally of attendance by private citizens at city commission and other local government meetings; another might maintain a record of the different topics addressed by private citizens at similar meetings; while a third might record the number of local government meeting attendees, letters, and phone calls received by the mayor and other public officials, and meetings held by special interest groups during a particular time period. As skilled researchers, we can readily see that each planner would be measuring (in a very simplistic fashion) a different dimension of citizen participation: extent of citizen participation, issues prompting citizen participation, and form of citizen participation. Therefore, the original *naming* of our variable, *citizen participation*, which was quite satisfactory from a conceptual point of view, proved inadequate for purposes of operationalization.

The precise and exact naming of variables is important in research. It is both essential to and a result of good operationalization. Variable names quite often evolve from an iterative process of forming a conceptual definition, then an operational definition, then renaming the concept to better match what can or will be measured. This looping process continues (our example illustrates only one iteration), resulting in a gradual refinement of the variable name and its measurement until a reasonable fit is obtained. Sometimes the concept of the variable that you end up with is a bit different from the original one that you started with, but at least you are measuring what you are talking about, if only because you are talking about what you are measuring!

the other end of the income spectrum. In studies of the general U.S. population, a bottom category of $5,000 or less usually works fine.

In studies of attitudes and orientations, the question of range of variation has another dimension. Unless you're careful, you may end up measuring only half an attitude without really meaning to. Here's an example of what I mean.

Suppose you're interested in people's attitudes toward expanding the use of nuclear power generators. You'd anticipate that some people consider nuclear power the greatest thing since the wheel, whereas other people have absolutely no interest in it. Given that anticipation, it would seem to make sense to ask people how much they favor expanding the use of nuclear energy and to give them answer categories ranging from "Favor it very much" to "Don't favor it at all."

This operationalization, however, conceals half the attitudinal spectrum regarding nuclear energy. Many people have feelings that go beyond simply not favoring it: They are, with greater or lesser degrees of intensity, actively opposed to it. In this instance, there is considerable variation on the left side of zero. Some oppose it a little, some quite a bit, and others a great deal. To measure the full range of variation, then, you'd want to operationalize attitudes toward nuclear energy with a range from favoring it very much, through no feelings one way or the other, to opposing it very much.

This consideration applies to many of the variables social scientists study. Virtually any public

issue involves both support and opposition, each in varying degrees. Political orientations range from very liberal to very conservative, and depending on the people you're studying, you may want to allow for radicals on one or both ends. Similarly, people are not just more or less religious; some are positively antireligious.

The point is not that you must measure the full range of variation in every case. You should, however, consider whether you need to, given your particular research purpose. If the difference between not religious and antireligious isn't relevant to your research, forget it. Someone has defined pragmatism as "any difference that makes no difference is no difference." Be pragmatic.

Finally, decisions on the range of variation should be governed by the expected distribution of attributes among the subjects of the study. In a study of college professors' attitudes toward the value of higher education, you could probably stop at no value and not worry about those who might consider higher education dangerous to students' health. (If you were studying students, however . . .)

Variations between the Extremes

Degree of precision is a second consideration in operationalizing variables. What it boils down to is how fine you will make distinctions among the various possible attributes composing a given variable. Does it matter for your purposes whether a person is 17 or 18 years old, or could you conduct your inquiry by throwing them together in a group labeled 10 to 19 years old? Don't answer too quickly. If you wanted to study rates of voter registration and participation, you'd definitely want to know whether the people you studied were old enough to vote. In general, if you're going to measure age, you must look at the purpose and procedures of your study and decide whether fine or gross differences in age are important to you. In a survey, you'll need to make these decisions in order to design an appropriate questionnaire. In the case of in-depth interviews, these decisions will condition the extent to which you probe for details.

The same thing applies to other variables. If you measure political affiliation, will it matter to your inquiry whether a person is a conservative Democrat rather than a liberal Democrat, or will it be sufficient to know the party? In measuring religious affiliation, is it enough to know that a person is Protestant, or do you need to know the denomination? Do you simply need to know whether or not a person is married, or will it make a difference to know if he or she has never married or is separated, widowed, or divorced?

There is, of course, no general answer to such questions. The answers come out of the purpose of a given study, or why we are making a particular measurement. I can give you a useful guideline, though. Whenever you're not sure how much detail to pursue in a measurement, get too much rather than too little. When a subject in an in-depth interview volunteers that she is 37 years old, record "37" in your notes, not "in her thirties." When you're analyzing the data, you can always combine precise attributes into more general categories, but you can never separate any variations you lumped together during observation and measurement.

A Note on Dimensions

We've already discussed dimensions as a characteristic of concepts. When researchers get down to the business of creating operational measures of variables, they often discover—or worse, never notice—that they're not exactly clear about which dimensions of a variable they're really interested in. Here's an example.

Let's suppose you're studying people's attitudes toward government, and you want to include an examination of how people feel about corruption. Here are just a few of the dimensions you might examine:

- Do people think there is corruption in government?
- How much corruption do they think there is?
- How certain are they in their judgment of how much corruption there is?
- How do they feel about corruption in government as a problem in society?
- What do they think causes it?

- Do they think it's inevitable?
- What do they feel should be done about it?
- What are they willing to do personally to eliminate corruption in government?
- How certain are they that they would be willing to do what they say they would do?

The list could go on and on—how people feel about corruption in government has many dimensions. It's essential to be clear about which ones are important in our inquiry; otherwise, you may measure how people *feel* about corruption when you really wanted to know how much they think there is, or vice versa.

Once you have determined how you're going to collect your data (for example, survey, field research) and have decided on the relevant range of variation, the degree of precision needed between the extremes of variation, and the specific dimensions of the variables that interest you, you may have another choice: a mathematical-logical one. That is, you may need to decide what level of measurement to use. To discuss this point, we need to take another look at attributes and their relationship to variables.

Defining Variables and Attributes

An attribute, you'll recall, is a characteristic or quality of something. *Female* is an example. So is *old* or *student*. Variables, on the other hand, are logical sets of attributes. Thus, *gender* is a variable composed of the attributes *female* and *male*.

The conceptualization and operationalization processes can be seen as the specification of variables and the attributes composing them. Thus, in the context of a study of unemployment, *employment status* is a variable having the attributes *employed* and *unemployed*; the list of attributes could also be expanded to include the other possibilities discussed earlier, such as *homemaker*.

Every variable must have two important qualities. First, the attributes composing it should be exhaustive. For the variable to have any utility in research, we must be able to classify every observation in terms of one of the attributes composing the variable. We'll run into trouble if we conceptualize the variable *political party affiliation* in terms of the attributes *Republican* and *Democrat*, because some of the people we set out to study will identify with the Green Party, the Reform Party, or some other organization, and some (often a large percentage) will tell us they have no party affiliation. We could make the list of attributes exhaustive by adding *other* and *no affiliation*. Whatever we do, we must be able to classify every observation.

At the same time, attributes composing a variable must be mutually exclusive. Every observation must be able to be classified in terms of one and only one attribute. For example, we need to define *employed* and *unemployed* in such a way that nobody can be both at the same time. That means being able to classify the person who is working at a job but is also looking for work. (We might run across a fully employed mud wrestler who is looking for the glamour and excitement of being a social researcher.) In this case, we might define the attributes so that *employed* takes precedence over *unemployed*, and anyone working at a job is employed regardless of whether he or she is looking for something better.

nominal measure A variable whose attributes have only the characteristics of exhaustiveness and mutual exclusiveness. In other words, a level of measurement describing a variable that has attributes that are merely different, as distinguished from *ordinal, interval,* or *ratio measures*. Gender is an example of a nominal measure.

Levels of Measurement

Attributes operationalized as mutually exclusive and exhaustive may be related in other ways as well. For example, the attributes composing variables may represent different levels of measurement. In this section, we'll examine four levels of measurement: nominal, ordinal, interval, and ratio.

Nominal Measures

Variables whose attributes have only the characteristics of exhaustiveness and mutual exclusiveness are **nominal measures**. Examples include *gender,*

religious affiliation, political party affiliation, birthplace, college major, and *hair color*. Although the attributes composing each of these variables—as *male* and *female* compose the variable *gender*—are distinct from one another (and exhaust the possibilities of gender among people), they have no additional structures. Nominal measures merely offer names or labels for characteristics.

Imagine a group of people characterized in terms of one such variable and physically grouped by the applicable attributes. For example, say we've asked a large gathering of people to stand together in groups according to the states in which they were born: all those born in Vermont in one group, those born in California in another, and so forth. The variable is *place of birth;* the attributes are *born in California, born in Vermont,* and so on. All the people standing in a given group have at least one thing in common and differ from the people in all other groups in that same regard. Where the individual groups form, how close they are to one another, or how the groups are arranged in the room is irrelevant. All that matters is that all the members of a given group share the same state of birth and that each group has a different shared state of birth. All we can say about two people in terms of a nominal variable is that they are either the same or different.

Ordinal Measures

Variables with attributes we can logically rank-order are **ordinal measures**. The different attributes of ordinal variables represent relatively more or less of the variable. Variables of this type are *social class, conservatism, alienation, prejudice, intellectual sophistication,* and the like. In addition to saying whether two people are the same or different in terms of an ordinal variable, you can also say one is "more" than the other—that is, more conservative, more religious, older, and so forth.

In the physical sciences, hardness is the most frequently cited example of an ordinal measure. We may say that one material (for example, diamond) is harder than another (say, glass) if the former can scratch the latter and not vice versa. By attempting to scratch various materials with other materials, we might eventually be able to arrange several materials in a row, ranging from the softest to the hardest. We could never say how hard a given material was in absolute terms; we could only say how hard in relative terms—which materials it is harder than and which softer than.

Let's pursue the earlier example of grouping the people at a social gathering. This time imagine that we ask all the people who have graduated from college to stand in one group, all those with only a high school diploma to stand in another group, and all those who have not graduated from high school to stand in a third group. This manner of grouping people satisfies the requirements for exhaustiveness and mutual exclusiveness discussed earlier. In addition, however, we might logically arrange the three groups in terms of the relative amount of formal education (the shared attribute) each had. We might arrange the three groups in a row, ranging from most to least formal education. This arrangement would provide a physical representation of an ordinal measure. If we knew which groups two individuals were in, we could determine that one had more, less, or the same formal education as the other.

Notice in this example that it is irrelevant how close or far apart the educational groups are from one another. The college and high school groups might be 5 feet apart, and the less-than-high-school group 500 feet farther down the line. These actual distances don't have any meaning. The high school group, however, should be between the less-than-high-school group and the college group, or else the rank order will be incorrect.

Interval Measures

For the attributes composing some variables, the actual distance separating those attributes does have meaning. Such variables are **interval measures**. For these, the logical distance between attributes can be expressed in meaningful standard intervals.

ordinal measure A level of measurement describing a variable with attributes we can rank-order along some dimension. An example is *socioeconomic status* as composed of the attributes *high, medium, low.*

For example, in the Fahrenheit temperature scale, the difference, or distance, between 80 degrees and 90 degrees is the same as that between 40 degrees and 50 degrees. However, 80 degrees Fahrenheit is not twice as hot as 40 degrees, because the zero point in the Fahrenheit scale is arbitrary; zero degrees does not really mean lack of heat. Similarly, minus 30 degrees on this scale doesn't represent 30 degrees less than no heat. (This is true for the Celsius scale as well. In contrast, the Kelvin scale is based on an absolute zero, which does mean a complete lack of heat.)

About the only interval measures commonly used in social scientific research are constructed measures such as standardized intelligence tests that have been more or less accepted. The interval separating IQ scores of 100 and 110 may be regarded as the same as the interval separating scores of 110 and 120 by virtue of the distribution of observed scores obtained by many thousands of people who have taken the tests over the years. But it would be incorrect to infer that someone with an IQ of 150 is 50 percent more intelligent than someone with an IQ of 100. (A person who received a score of 0 on a standard IQ test could not be regarded, strictly speaking, as having no intelligence, although we might feel he or she was unsuited to be a college professor or even a college student. But perhaps a dean . . . ?)

When comparing two people in terms of an interval variable, we can say they are different from one another (nominal), and that one is more than another (ordinal). In addition, we can say "how much" more.

interval measure A level of measurement describing a variable whose attributes are rank-ordered and have equal distances between adjacent attributes. The Fahrenheit temperature scale is an example of this, because the distance between 17 and 18 is the same as that between 89 and 90.

ratio measure A level of measurement describing a variable with attributes that have all the qualities of nominal, ordinal, and interval measures and in addition are based on a "true zero" point. Age is an example of a ratio measure.

Ratio Measures

Most of the social scientific variables meeting the minimum requirements for interval measures also meet the requirements for ratio measures. In **ratio measures**, the attributes composing a variable, besides having all the structural characteristics mentioned previously, are based on a true zero point. The Kelvin temperature scale is one such measure. Examples from social scientific research include age, length of residence in a given place, number of organizations belonged to, number of times attending religious services during a particular period of time, number of times married, and number of Arab friends.

Returning to the illustration of methodological party games, we might ask a gathering of people to group themselves by age. All the one-year-olds would stand (or sit or lie) together, the two-year-olds together, the three-year-olds, and so forth. The fact that members of a single group share the same age and that each different group has a different shared age satisfies the minimum requirements for a nominal measure. Arranging the several groups in a line from youngest to oldest meets the additional requirements of an ordinal measure and lets us determine if one person is older than, younger than, or the same age as another. If we space the groups equally far apart, we satisfy the additional requirements of an interval measure and can say how much older one person is than another. Finally, because one of the attributes included in age represents a true zero (babies carried by women about to give birth), the phalanx of hapless party goers also meets the requirements of a ratio measure, permitting us to say that one person is twice as old as another. (Remember this in case you're asked about it in a workbook assignment.) Another example of a ratio measure is income, which extends from an absolute zero to approximately infinity, if you happen to be the founder of Microsoft.

Comparing two people in terms of a ratio variable, then, allows us to conclude (1) whether they

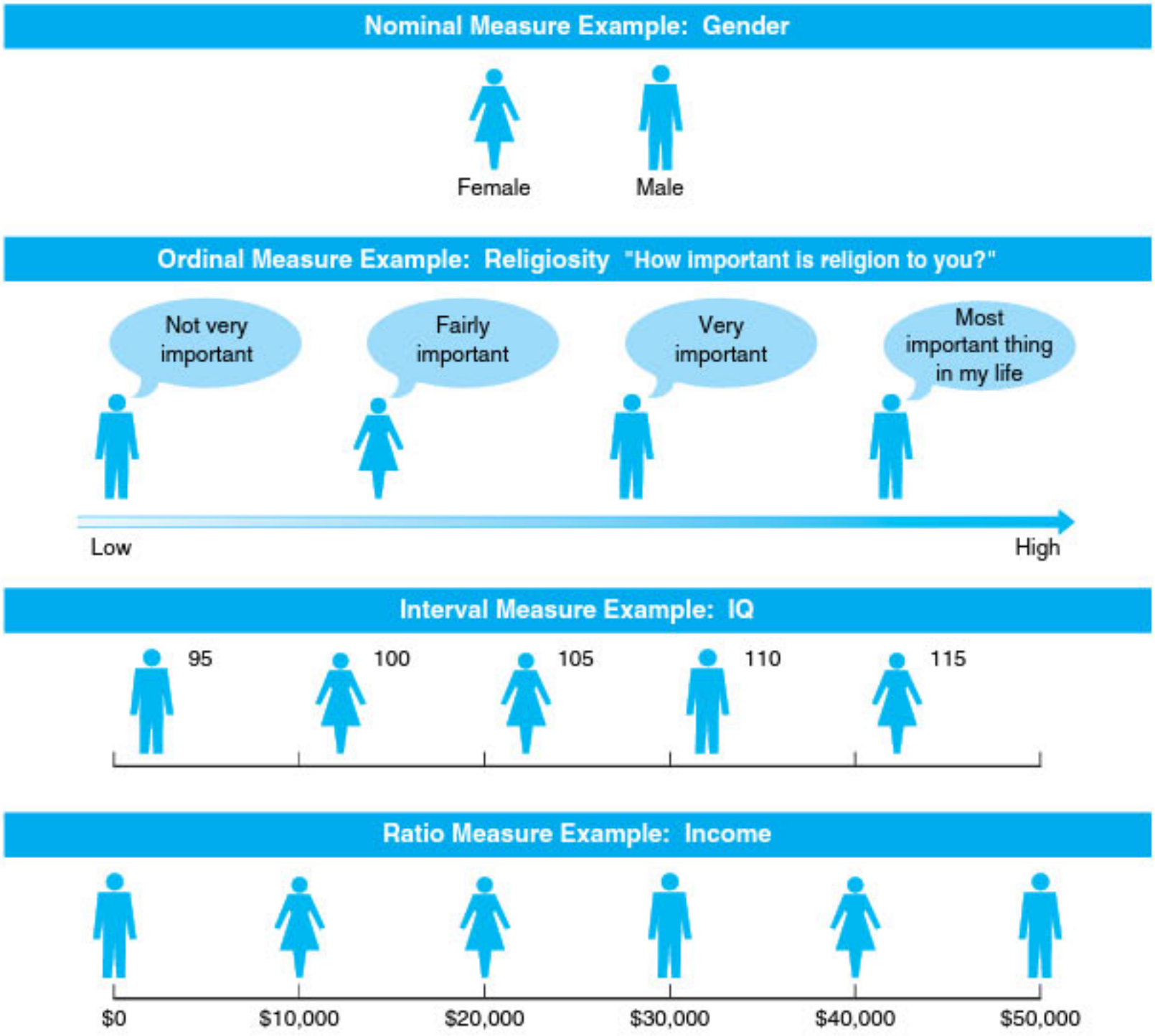


FIGURE 5-1

Levels of Measurement. Often you can choose among different levels of measurement—nominal, ordinal, interval, or ratio—carrying progressively more amounts of information.

are different (or the same), (2) whether one is more than the other, (3) how much they differ, and (4) what the ratio of one to another is. Figure 5-1 summarizes this discussion by presenting a graphic illustration of the four levels of measurement.

Implications 1of Levels of Measurement

Because it's unlikely that you'll undertake the physical grouping of people just described (try it once, and you won't be invited to many parties), I should draw your attention to some of the practical implications of the differences that have been distinguished. These implications appear primarily in the analysis of data (discussed in Part 4), but you need to anticipate such implications when you're structuring any research project.

Certain quantitative analysis techniques require variables that meet certain minimum levels of

measurement. To the extent that the variables to be examined in a research project are limited to a particular level of measurement—say, ordinal—you should plan your analytical techniques accordingly. More precisely, you should anticipate drawing research conclusions appropriate to the levels of measurement used in your variables. For example, you might reasonably plan to determine and report the mean age of a population under study (add up all the individual ages and divide by the number of people), but you should not plan to report the mean religious affiliation, because that is a nominal variable, and the mean requires ratio-level data. (You could report the modal—the most common—religious affiliation.)

At the same time, you can treat some variables as representing different levels of measurement. Ratio measures are the highest level, descending through interval and ordinal to nominal, the lowest level of measurement. A variable representing a higher level of measurement—say, ratio—can also be treated as representing a lower level of measurement—say, ordinal. Recall, for example, that age is a ratio measure. If you wished to examine only the relationship between age and some ordinal-level variable—say, self-perceived religiosity: high, medium, and low—you might choose to treat age as an ordinal-level variable as well. You might characterize the subjects of your study as being young, middle-aged, and old, specifying what age range composed each of these groupings. Finally, age might be used as a nominal-level variable for certain research purposes. People might be grouped as being born during the Depression or not. Another nominal measurement, based on birth date rather than just age, would be the grouping of people by astrological signs.

The level of measurement you'll seek, then, is determined by the analytical uses you've planned for a given variable, keeping in mind that some variables are inherently limited to a certain level. If a variable is to be used in a variety of ways, requiring different levels of measurement, the study should be designed to achieve the highest level required. For example, if the subjects in a study are asked their exact ages, they can later be organized into ordinal or nominal groupings.

Again, you need not necessarily measure variables at their highest level of measurement. If you're sure to have no need for ages of people at higher than the ordinal level of measurement, you may simply ask people to indicate their age range, such as 20 to 29, 30 to 39, and so forth. In a study of the wealth of corporations, rather than seek more precise information, you may use Dun & Bradstreet ratings to rank corporations. Whenever your research purposes are not altogether clear, however, seek the highest level of measurement possible. As we've discussed, although ratio measures can later be reduced to ordinal ones, you cannot convert an ordinal measure to a ratio one. More generally, you cannot convert a lower-level measure to a higher-level one. That is a one-way street worth remembering.

Typically a research project will tap variables at different levels of measurement. For example, William Bielby and Denise Bielby (1999) set out to examine the world of film and television, using a nomothetic, longitudinal approach (take a moment to remind yourself what that means). In what they referred to as the "culture industry," the authors found that *reputation* (an ordinal variable) is the best predictor of screenwriters' future productivity. More interestingly, they found that screenwriters who were represented by "core" (or elite) agencies were not only far more likely to find jobs (a nominal variable), but also jobs that paid more (a ratio variable). In other words, the researchers found that agencies' reputations (ordinal) was a key independent variable for predicting a screenwriter's career success. The researchers also found that being older (ratio), female (nominal), an ethnic minority (nominal), and having more years of experience (ratio) were disadvantageous for a writer's career. On the other hand, higher earnings from previous years (measured in ordinal categories) led to more success in the future. In Bielby and Bielby's terms, "success breeds success" (1999: 80).

Single or Multiple Indicators

With so many alternatives for operationalizing social scientific variables, you may find yourself worrying about making the right choices. To

counter this feeling, let me add a momentary dash of certainty and stability.

Many social research variables have fairly obvious, straightforward measures. No matter how you cut it, gender usually turns out to be a matter of male or female: a nominal-level variable that can be measured by a single observation—either by looking (well, not always) or by asking a question (usually). In a study involving the size of families, you'll want to think about adopted and foster children, as well as blended families, but it's usually pretty easy to find out how many children a family has. For most research purposes, the resident population of a country is the resident population of that country—you can look it up in an almanac and know the answer. A great many variables, then, have obvious single indicators. If you can get one piece of information, you have what you need.

Sometimes, however, there is no single indicator that will give you the measure of a variable you really want. As discussed earlier in this chapter, many concepts are subject to varying interpretations—each with several possible indicators. In these cases, you'll want to make several observations for a given variable. You can then combine the several pieces of information you've collected, creating a composite measurement of the variable in question. Chapter 6 is devoted to ways of doing that, so here let's just discuss one simple illustration.

Consider the concept "college performance." All of us have noticed that some students perform well in college courses and others don't. In studying these differences, we might ask what characteristics and experiences are related to high levels of performance (many researchers have done just that). How should we measure overall performance? Each grade in any single course is a potential indicator of college performance, but it also may not typify the student's general performance. The solution to this problem is so firmly established that it is, of course, obvious: the grade point average (GPA). We assign numerical scores to each letter grade, total the points earned by a given student, and divide by the number of courses taken, thus obtaining a composite measure. (If the courses vary in number of credits, we adjust the point values accordingly.) Creating such composite measures in social research is often appropriate.

Some Illustrations of Operationalization Choices

To bring together all the operationalization choices available to the social researcher and to show the potential in those possibilities, let's look at some of the distinct ways you might address various research problems. The alternative ways of operationalizing the variables in each case should demonstrate the opportunities that social research can present to our ingenuity and imaginations. To simplify matters, I have not attempted to describe all the research conditions that would make one alternative superior to the others, though in a given situation they would not all be equally appropriate.

Here are specific research questions, then, and some of the ways you could address them. We'll begin with an example discussed earlier in the chapter. It has the added advantage that one of the variables is straightforward to operationalize.

1. Are women more compassionate than men?
   a. Select a group of subjects for study, with equal numbers of men and women. Present them with hypothetical situations that involve someone's being in trouble. Ask them what they would do if they were confronted with that situation. What would they do, for example, if they came across a small child who was lost and crying for his or her parents? Consider any answer that involves helping or comforting the child as an indicator of compassion. See whether men or women are more likely to indicate they would be compassionate.
   b. Set up an experiment in which you pay a small child to pretend that he or she is lost. Put the child to work on a busy sidewalk and observe whether men or women are more likely to offer assistance. Also be sure to count the total number of men and

women who walk by, because there may be more of one than the other. If that's the case, simply calculate the percentage of men and the percentage of women who help.

c. Select a sample of people and do a survey in which you ask them what organizations they belong to. Calculate whether women or men are more likely to belong to those that seem to reflect compassionate feelings. To take account of men who belong to more organizations than do women in general—or vice versa—do this: For each person you study, calculate the percentage of his or her organizational memberships that reflect compassion. See if men or women have a higher average percentage.

2. Are sociology students or accounting students better informed about world affairs?
   a. Prepare a short quiz on world affairs and arrange to administer it to the students in a sociology class and in an accounting class at a comparable level. If you want to compare sociology and accounting majors, be sure to ask students what they are majoring in.
   b. Get the instructor of a course in world affairs to give you the average grades of sociology and accounting students in the course.
   c. Take a petition to sociology and accounting classes that urges that "the United Nations headquarters be moved to New York City." Keep a count of how many in each class sign the petition and how many inform you that the UN headquarters is already located in New York City.
3. Do people consider New York or California the better place to live?
   a. Consulting the *Statistical Abstract of the United States* or a similar publication, check the migration rates into and out of each state. See if you can find the numbers moving directly from New York to California and vice versa.
   b. The national polling companies—Gallup, Harris, Roper, and so forth—often ask people what they consider the best state to live in. Look up some recent results in the library or through your local newspaper.
   c. Compare suicide rates in the two states.
4. Who are the most popular instructors on your campus, those in the social sciences, the natural sciences, or the humanities?
   a. If your school has a provision for student evaluation of instructors, review some recent results and compute the average rating of each of the three groups.
   b. Begin visiting the introductory courses given in each group of disciplines and measure the attendance rate of each class.
   c. In December, select a group of faculty in each of the three divisions and ask them to keep a record of the numbers of holiday greeting cards and presents they receive from admiring students. See who wins.

The point of these examples is not necessarily to suggest respectable research projects but to illustrate the many ways variables can be operationalized.

Operationalization Goes On and On

Although I've discussed conceptualization and operationalization as activities that precede data collection and analysis—for example, you must design questionnaire items before you send out a questionnaire—these two processes continue throughout any research project, even if the data have been collected in a structured mass survey. As we've seen, in less-structured methods such as field research, the identification and specification of relevant concepts is inseparable from the ongoing process of observation.

As a researcher, always be open to reexamining your concepts and definitions. The ultimate purpose of social research is to clarify the nature of social life. The validity and utility of what you learn in this regard doesn't depend on when you first figured out how to look at things any more than it matters whether you got the idea from a learned textbook, a dream, or your brother-in-law.

Criteria of Measurement Quality

This chapter has come some distance. It began with the bald assertion that social scientists can measure anything that exists. Then we discovered that most of the things we might want to measure and study don't really exist. Next we learned that it's possible to measure them anyway. Now we conclude the chapter with a discussion of some of the yardsticks against which we judge our relative success or failure in measuring things—even things that don't exist.

Precision and Accuracy

To begin, measurements can be made with varying degrees of precision. As we saw in the discussion of operationalization, precision concerns the fineness of distinctions made between the attributes that compose a variable. The description of a woman as "43 years old" is more precise than "in her forties." Saying a street-corner gang was formed "in the summer of 1996" is more precise than saying "during the 1990s."

As a general rule, precise measurements are superior to imprecise ones, as common sense dictates. There are no conditions under which imprecise measurements are intrinsically superior to precise ones. Even so, exact precision is not always necessary or desirable. If knowing that a woman is in her forties satisfies your research requirements, then any additional effort invested in learning her precise age is wasted. The operationalization of concepts, then, must be guided partly by an understanding of the degree of precision required. If your needs are not clear, be more precise rather than less.

Don't confuse precision with accuracy, however. Describing someone as "born in New England" is less precise than "born in Stowe, Vermont"—but suppose the person in question was actually born in Boston. The less-precise description, in this instance, is more accurate, a better reflection of the real world.

Precision and accuracy are obviously important qualities in research measurement, and they probably need no further explanation. When social scientists construct and evaluate measurements, however, they pay special attention to two technical considerations: reliability and validity.

Reliability

In the abstract, **reliability** is a matter of whether a particular technique, applied repeatedly to the same object, yields the same result each time. Let's say you want to know how much I weigh. (No, I don't know why.) As one technique, say you ask two different people to estimate my weight. If the first person estimates 150 pounds and the other estimates 300, we have to conclude the technique of having people estimate my weight isn't very reliable.

Suppose, as an alternative, that you use a bathroom scale as your measurement technique. I step on the scale twice, and you note the same result each time. The scale has presumably reported the same weight for me both times, indicating that the scale provides a more reliable technique for measuring a person's weight than asking people to estimate it does.

Reliability, however, does not ensure accuracy any more than precision does. Suppose I've set my bathroom scale to shave five pounds off my weight just to make me feel better. Although you would (reliably) report the same weight for me each time, you would always be wrong. This new element, called *bias*, is discussed in Chapter 8. For now, just be warned that reliability does not ensure accuracy.

Let's suppose we're interested in studying morale among factory workers in two different kinds

reliability That quality of measurement method that suggests that the same data would have been collected each time in repeated observations of the same phenomenon. In the context of a survey, we would expect that the question "Did you attend religious services last week?" would have higher reliability than the question "About how many times have you attended religious services in your life?" This is not to be confused with *validity*.

of factories. In one set of factories, workers have specialized jobs, reflecting an extreme division of labor. Each worker contributes a tiny part to the overall process performed on a long assembly line. In the other set of factories, each worker performs many tasks, and small teams of workers complete the whole process.

How should we measure morale? Following one strategy, we could observe the workers in each factory, noticing such things as whether they joke with one another, whether they smile and laugh a lot, and so forth. We could ask them how they like their work and even ask them whether they think they would prefer their current arrangement or the other one being studied. By comparing what we observed in the different factories, we might reach a conclusion about which assembly process produces the higher morale. Notice that I've just described a qualitative measurement procedure.

Now let's look at some reliability problems inherent in this method. First, how you and I are feeling when we do the observing will likely color what we see. We may misinterpret what we see. We may see workers kidding each other but think they're having an argument. We may catch them on an off day. If we were to observe the same group of workers several days in a row, we might arrive at different evaluations on each day. Further, even if several observers evaluated the same behavior, they might arrive at different conclusions about the workers' morale.

Here's another strategy for assessing morale, a quantitative approach. Suppose we check the company records to see how many grievances have been filed with the union during some fixed period. Presumably this would be an indicator of morale: the more grievances, the lower the morale. This measurement strategy would appear to be more reliable: Counting up the grievances over and over, we should keep arriving at the same number.

If you find yourself thinking that the number of grievances doesn't necessarily measure morale, you're worrying about validity, not reliability. We'll discuss validity in a moment. The point for now is that the last method is more like my bathroom scale—it gives consistent results.

In social research, reliability problems crop up in many forms. Reliability is a concern every time a single observer is the source of data, because we have no certain guard against the impact of that observer's subjectivity. We can't tell for sure how much of what's reported originated in the situation observed and how much in the observer.

Subjectivity is not only a problem with single observers, however. Survey researchers have known for a long time that different interviewers, because of their own attitudes and demeanors, get different answers from respondents. Or, if we were to conduct a study of newspapers' editorial positions on some public issue, we might create a team of coders to take on the job of reading hundreds of editorials and classifying them in terms of their position on the issue. Unfortunately, different coders will code the same editorial differently. Or we might want to classify a few hundred specific occupations in terms of some standard coding scheme, say a set of categories created by the Department of Labor or by the Census Bureau. You and I would not place all those occupations in the same categories.

Each of these examples illustrates problems of reliability. Similar problems arise whenever we ask people to give us information about themselves. Sometimes we ask questions that people don't know the answers to: How many times have you been to religious services? Sometimes we ask people about things they consider totally irrelevant: Are you satisfied with China's current relationship with Albania? In such cases, people will answer differently at different times because they're making up answers as they go. Sometimes we explore issues so complicated that a person who had a clear opinion in the matter might arrive at a different interpretation of the question when asked a second time.

So how do you create reliable measures? If your research design calls for asking people for information, you can be careful to ask only about things the respondents are likely to know the answer to. Ask about things relevant to them, and be clear in what you're asking. Of course, these techniques don't solve every possible reliability problem. Fortunately, social researchers have developed

several techniques for cross-checking the reliability of the measures they devise.

Test-Retest Method

Sometimes it's appropriate to make the same measurement more than once, a technique called the *test-retest method*. If you don't expect the sought-after information to change, then you should expect the same response both times. If answers vary, the measurement method may, to the extent of that variation, be unreliable. Here's an illustration.

In their research on Health Hazard Appraisal (HHA), a part of preventive medicine, Jeffrey Sacks, W. Mark Krushat, and Jeffrey Newman (1980) wanted to determine the risks associated with various background and lifestyle factors, making it possible for physicians to counsel their patients appropriately. By knowing patients' life situations, physicians could advise them on their potential for survival and on how to improve it. This purpose, of course, depended heavily on the accuracy of the information gathered about each subject in the study.

To test the reliability of their information, Sacks and his colleagues had all 207 subjects complete a baseline questionnaire that asked about their characteristics and behavior. Three months later, a follow-up questionnaire asked the same subjects for the same information, and the results of the two surveys were compared. Overall, only 15 percent of the subjects reported the same information in both studies.

Sacks and his colleagues report the following:

> Almost 10 percent of subjects reported a different height at follow-up examination. Parental age was changed by over one in three subjects. One parent reportedly aged 20 chronologic years in three months. One in five ex-smokers and ex-drinkers have apparent difficulty in reliably recalling their previous consumption pattern.
>
> *(1980: 730)*

Some subjects erased all trace of previously reported heart murmur, diabetes, emphysema, arrest record, and thoughts of suicide. One subject's mother, deceased in the first questionnaire, was apparently alive and well in time for the second. One subject had one ovary missing in the first study but present in the second. In another case, an ovary present in the first study was missing in the second study—and had been for ten years! One subject was reportedly 55 years old in the first study and 50 years old three months later. (You have to wonder whether the physician-counselors could ever have nearly the impact on their patients that their patients' memories did.) Thus, test-retest revealed that this data-collection method was not especially reliable.

Split-Half Method

As a general rule, it's always good to make more than one measurement of any subtle or complex social concept, such as prejudice, alienation, or social class. This procedure lays the groundwork for another check on reliability. Let's say you've created a questionnaire that contains ten items you believe measure prejudice against women. Using the split-half technique, you would randomly assign those ten items to two sets of five. Each set should provide a good measure of prejudice against women, and the two sets should classify respondents the same way. If the two sets of items classify people differently, you most likely have a problem of reliability in your measure of the variable.

Using Established Measures

Another way to help ensure reliability in getting information from people is to use measures that have proved their reliability in previous research. If you want to measure anomia, for example, you might want to follow Srole's lead.

The heavy use of measures, though, does not guarantee their reliability. For example, the Scholastic Assessment Tests (SATs) and the Minnesota Multiphasic Personality Inventory (MMPI) have been accepted as established standards in their respective domains for decades. In recent years, though, they've needed fundamental overhauling to reflect changes in society, eliminating outdated topics and gender bias in wording.

Reliability of Research Workers

As we've seen, it's also possible for measurement unreliability to be generated by research workers: interviewers and coders, for example. There are several ways to check on reliability in such cases. To guard against interviewer unreliability in surveys, for example, a supervisor will call a subsample of the respondents on the telephone and verify selected pieces of information.

Replication works in other situations also. If you're worried that newspaper editorials or occupations may not be classified reliably, you could have each independently coded by several coders. Those cases that are classified inconsistently can then be evaluated more carefully and resolved.

Finally, clarity, specificity, training, and practice can prevent a great deal of unreliability and grief. If you and I spent some time reaching a clear agreement on how to evaluate editorial positions on an issue—discussing various positions and reading through several together—we could probably do a good job of classifying them in the same way independently.

The reliability of measurements is a fundamental issue in social research, and we'll return to it more than once in the chapters ahead. For now, however, let's recall that even total reliability doesn't ensure that our measures actually measure what we think they measure. Now let's plunge into the question of validity.

validity A term describing a measure that accurately reflects the concept it is intended to measure. For example, your IQ would seem a more valid measure of your intelligence than the number of hours you spend in the library would. Though the ultimate validity of a measure can never be proved, we may agree to its relative validity on the basis of face validity, criterion validity, content validity, construct validity, internal validation, and external validation. This must not be confused with *reliability*.

face validity That quality of an indicator that makes it seem a reasonable measure of some variable. That the frequency of attendance at religious services is some indication of a person's religiosity seems to make sense without a lot of explanation. It has face validity.

Validity

In conventional usage, **validity** refers to the extent to which an empirical measure adequately reflects the real meaning of the concept under consideration. Whoops! I've already committed us to the view that concepts don't have real meanings. How can we ever say whether a particular measure adequately reflects the concept's meaning, then? Ultimately, of course, we can't. At the same time, as we've already seen, all of social life, including social research, operates on agreements about the terms we use and the concepts they represent. There are several criteria of success in making measurements that are appropriate to these agreed-on meanings of concepts.

First, there's something called **face validity**. Particular empirical measures may or may not jibe with our common agreements and our individual mental images concerning a particular concept. For example, you and I might quarrel about whether counting the number of grievances filed with the union will adequately measure morale. Still, we'd surely agree that the number of grievances has *something* to do with morale. That is, the measure is valid "on its face," whether or not it's adequate. If I were to suggest that we measure morale by finding out how many books the workers took out of the library during their off-duty hours, you'd undoubtedly raise a more serious objection: That measure wouldn't have much face validity.

Second, I've already pointed to many of the more formally established agreements that define some concepts. The Census Bureau, for example, has created operational definitions of such concepts as family, household, and employment status that seem to have a workable validity in most studies using these concepts.

Three additional types of validity also specify particular ways of testing the validity of measures. The first, **criterion-related validity**, sometimes called *predictive validity*, is based on some external criterion. For example, the validity of College Board exams is shown in their ability to predict students'

success in college. The validity of a written driver's test is determined, in this sense, by the relationship between the scores people get on the test and their subsequent driving records. In these examples, college success and driving ability are the criteria.

To test your understanding of criterion-related validity, see whether you can think of behaviors that might be used to validate each of the following attitudes:

Is very religious

Supports equality of men and women

Supports far-right militia groups

Is concerned about the environment

Some possible validators would be, respectively, attends religious services, votes for women candidates, belongs to the NRA, and belongs to the Sierra Club.

Sometimes it's difficult to find behavioral criteria that can be taken to validate measures as directly as in such examples. In those instances, however, we can often approximate such criteria by applying a different test. We can consider how the variable in question ought, theoretically, to relate to other variables. **Construct validity** is based on the logical relationships among variables.

Suppose, for example, that you want to study the sources and consequences of marital satisfaction. As part of your research, you develop a measure of marital satisfaction, and you want to assess its validity.

In addition to developing your measure, you'll have developed certain theoretical expectations about the way the variable *marital satisfaction* relates to other variables. For example, you might reasonably conclude that satisfied husbands and wives will be less likely than dissatisfied ones to cheat on their spouses. If your measure relates to marital fidelity in the expected fashion, that constitutes evidence of your measure's construct validity. If satisfied marriage partners are as likely to cheat on their spouses as are the dissatisfied ones, however, that would challenge the validity of your measure.

Tests of construct validity, then, can offer a weight of evidence that your measure either does or doesn't tap the quality you want it to measure, without providing definitive proof. Although I have suggested that tests of construct validity are less compelling than those of criterion validity, there is room for disagreement about which kind of test a particular comparison variable *(driving record, marital fidelity)* represents in a given situation. It's less important to distinguish the two types of validity tests than to understand the logic of validation that they have in common: If we've succeeded in measuring some variable, then our measures should relate in some logical way to other measures.

Finally, **content validity** refers to how much a measure covers the range of meanings included within a concept. For example, a test of mathematical ability cannot be limited to addition but also needs to cover subtraction, multiplication, division, and so forth. Or, if we're measuring prejudice, do our measurements reflect all types of prejudice, including prejudice against racial and ethnic groups, religious minorities, women, the elderly, and so on?

Figure 5-2 presents a graphic portrayal of the difference between validity and reliability. If you think of measurement as analogous to repeatedly shooting at the bull's-eye on a target, you'll see that reliability looks like a "tight pattern," regardless of where the shots hit, because reliability is a function of consistency. Validity, on the other hand, is a function of shots being arranged around the bull's-eye. The failure of reliability in the figure is randomly distributed around the target; the failure of validity is systematically off the mark. Notice that neither an unreliable nor an invalid measure is likely to be very useful.

criterion-related validity The degree to which a measure relates to some external criterion. For example, the validity of College Board tests is shown in their ability to predict the college success of students. Also called *predictive validity*.

construct validity The degree to which a measure relates to other variables as expected within a system of theoretical relationships.

content validity The degree to which a measure covers the range of meanings included within a concept.

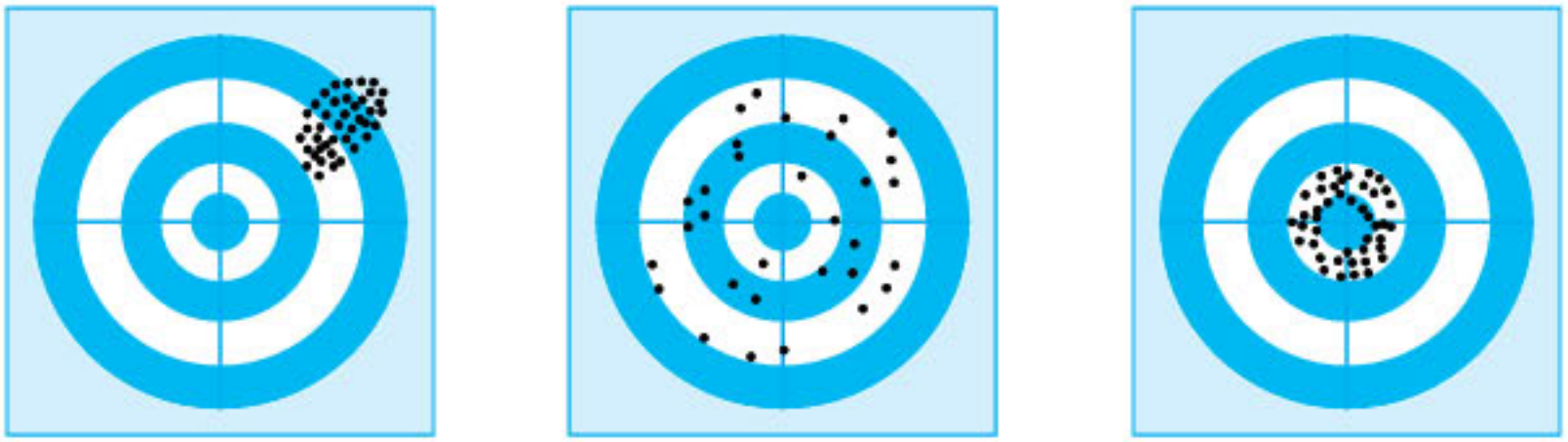


FIGURE 5-2

An Analogy to Validity and Reliability. A good measurement technique should be both valid (measuring what it is intended to measure) and reliable (yielding a given measurement dependably).

Who Decides What's Valid?

Our discussion of validity began with a reminder that we depend on agreements to determine what's real, and we've just seen some of the ways social scientists can agree among themselves that they have made valid measurements. There is yet another way of looking at validity.

Social researchers sometimes criticize themselves and one another for implicitly assuming they are somewhat superior to those they study. For example, researchers often seek to uncover motivations that the social actors themselves are unaware of. You think you bought that new Burpo-Blasto because of its high performance and good looks, but *we* know you're really trying to achieve a higher social status.

This implicit sense of superiority would fit comfortably with a totally positivistic approach (the biologist feels superior to the frog on the lab table), but it clashes with the more humanistic and typically qualitative approach taken by many social scientists. We'll explore this issue more deeply in Chapter 10.

In seeking to understand the way ordinary people make sense of their worlds, ethnomethodologists have urged all social scientists to pay more respect to these natural social processes of conceptualization and shared meaning. At the very least, behavior that may seem irrational from the scientist's paradigm may make logical sense when viewed through the actor's paradigm.

Ultimately, social researchers should look both to their colleagues and to their subjects as sources of agreement on the most useful meanings and measurements of the concepts they study. Sometimes one source will be more useful, sometimes the other. But neither one should be dismissed.

Tension between Reliability and Validity

Clearly, we want our measures to be both reliable and valid. However, a tension often arises between the criteria of reliability and validity, forcing a trade-off between the two.

Recall the example of measuring morale in different factories. The strategy of immersing yourself in the day-to-day routine of the assembly line, observing what goes on, and talking to the workers would seem to provide a more valid measure of morale than counting grievances would. It just seems obvious that we'd get a clearer sense of whether the morale was high or low using this first method.

As I pointed out earlier, however, the counting strategy would be more reliable. This situation reflects a more general strain in research measurement. Most of the really interesting concepts we want to study have many subtle nuances, so specifying precisely what we mean by them is hard. Researchers sometimes speak of such concepts as having a "richness of meaning." Although scores of books and articles have been written on the topic

of anomie/anomia, for example, they still haven't exhausted its meaning.

Very often, then, specifying reliable operational definitions and measurements seems to rob concepts of their richness of meaning. Positive morale is much more than a lack of grievances filed with the union; anomia is much more than what is measured by the five items created by Leo Srole. Yet, the more variation and richness we allow for a concept, the more opportunity there is for disagreement on how it applies to a particular situation, thus reducing reliability.

To some extent, this dilemma explains the persistence of two quite different approaches to social research: quantitative, nomothetic, structured techniques such as surveys and experiments on the one hand, and qualitative, idiographic methods such as field research and historical studies on the other. In the simplest generalization, the former methods tend to be more reliable, the latter more valid.

By being forewarned, you'll be effectively forearmed against this persistent and inevitable dilemma. If there is no clear agreement on how to measure a concept, measure it several different ways. If the concept has several dimensions, measure them all. Above all, know that the concept does not have any meaning other than what you and I give it. The only justification for giving any concept a particular meaning is utility. Measure concepts in ways that help us understand the world around us.

MAIN POINTS

Introduction

- The interrelated processes of conceptualization, operationalization, and measurement allow researchers to move from a general idea about what they want to study to effective and well-defined measurements in the real world.

Measuring Anything That Exists

- Conceptions are mental images we use as summary devices for bringing together observations and experiences that seem to have something in common. We use terms or labels to reference these conceptions.
- Concepts are constructs; they represent the agreed-on meanings we assign to terms. Our concepts don't exist in the real world, so they can't be measured directly, but we can measure the things that our concepts summarize.

Conceptualization

- Conceptualization is the process of specifying observations and measurements that give concepts definite meaning for the purposes of a research study.
- Conceptualization includes specifying the indicators of a concept and describing its dimensions. Operational definitions specify how variables relevant to a concept will be measured.

Definitions in Descriptive and Explanatory Studies

- Precise definitions are even more important in descriptive than in explanatory studies. The degree of precision needed varies with the type and purpose of a study.

Operationalization Choices

- Operationalization is an extension of conceptualization that specifies the exact procedures that will be used to measure the attributes of variables.
- Operationalization involves a series of interrelated choices: specifying the range of variation that is appropriate for the purposes of a study, determining how precisely to measure variables, accounting for relevant dimensions of variables, clearly defining the attributes of variables and their relationships, and deciding on an appropriate level of measurement.
- Researchers must choose from four levels of measurement, which capture increasing amounts of information: nominal, ordinal, interval, and ratio. The most appropriate level depends on the purpose of the measurement.
- A given variable can sometimes be measured at different levels. When in doubt, researchers should use the highest level of measurement appropriate to that variable so they can capture the greatest amount of information.

- Operationalization begins in the design phase of a study and continues through all phases of the research project, including the analysis of data.

Criteria of Measurement Quality

- Criteria of the quality of measures include precision, accuracy, reliability, and validity.
- Whereas reliability means getting consistent results from the same measure, validity refers to getting results that accurately reflect the concept being measured.
- Researchers can test or improve the reliability of measures through the test-retest method, the split-half method, the use of established measures, and the examination of work performed by research workers.
- The yardsticks for assessing a measure's validity include face validity, criterion-related validity, construct validity, and content validity.
- Creating specific, reliable measures often seems to diminish the richness of meaning our general concepts have. This problem is inevitable. The best solution is to use several different measures, tapping the different aspects of a concept.

KEY TERMS

The following terms are defined in context in the chapter and at the bottom of the page where the term is introduced, as well as in the comprehensive glossary at the back of the book.

conceptualization
construct validity
content validity
criterion-related validity
dimension
face validity
indicator
interval measure
nominal measure
ordinal measure
ratio measure
reliability
specification
validity

REVIEW QUESTIONS AND EXERCISES

1. Pick a social science concept such as liberalism or alienation, then specify that concept so that it could be studied in a research project. Be sure to specify the indicators you'll use as well as the dimensions you wish to include in and exclude from your conceptualization.
2. What level of measurement—nominal, ordinal, interval, or ratio—describes each of the following variables?
   a. Race (white, African American, Asian, and so on)
   b. Order of finish in a race (first, second, third, and so on)
   c. Number of children in families
   d. Populations of nations
   e. Attitudes toward nuclear energy (strongly approve, approve, disapprove, strongly disapprove)
   f. Region of birth (Northeast, Midwest, and so on)
   g. Political orientation (very liberal, somewhat liberal, somewhat conservative, very conservative)
3. Let's conceptualize the variable: prejudice. Using your favorite web browser, search for the term *prejudice*. After reviewing several of the websites resulting from your search, make a list of some different forms of prejudice that might be studied in an omnibus project dealing with that topic.
4. Let's discover *truth*. In a good dictionary, look up *truth* and *true*, then copy out the definitions. Note the key terms used in those definitions (e.g., *reality*), look up the definitions of those terms, and copy out these definitions as well. Continue this process until no new terms appear. Comment on what you've learned from this exercise.

ADDITIONAL READINGS

Bohrnstedt, George W. 1983. "Measurement." Pp. 70–121 in *Handbook of Survey Research*, edited by Peter H. Rossi, James D. Wright, and Andy B. Anderson. New York: Academic Press. This essay offers the logical and statistical grounding of reliability and validity in measurement.

Grimes, Michael D. 1991. *Class in Twentieth-Century American Sociology: An Analysis of Theories and Measurement Strategies*. New York: Praeger. This book provides an excellent, long-term view of conceptualization as the author examines a variety of theoretical views of social class and the measurement techniques appropriate to those theories.

Lazarsfeld, Paul F., and Morris Rosenberg, eds. 1955. *The Language of Social Research*, Section I. New York: Free Press of Glencoe. An excellent and diverse classic collection of descriptions of specific measurements in past social research. These 14 articles present useful and readable accounts of actual measurement operations performed by social researchers, as well as more conceptual discussions of measurement in general.

Miller, Delbert. 1991. *Handbook of Research Design and Social Measurement*. Newbury Park, CA: Sage. A powerful reference work. This book, especially Part 6, cites and describes a wide variety of operational measures used in earlier social research. In several cases, the questionnaire formats used are presented. Though the quality of these illustrations is uneven, they provide excellent examples of possible variations.

Silverman, David. 1993. *Interpreting Qualitative Data: Methods for Analyzing Talk, Text, and Interaction*, Chapter 7. Newbury Park, CA: Sage. This chapter deals with the issues of validity and reliability specifically in regard to qualitative research.

U.S. Department of Health and Human Services. 1992. *Survey Measurement of Drug Use*. Washington, DC: U.S. Government Printing Office. An extensive review of techniques devised and used for measuring various kinds of drug use.

SPSS EXERCISES

See the booklet that accompanies your text for exercises using SPSS (Statistical Package for the Social Sciences). There are exercises offered for each chapter, and you'll also find a detailed primer on using SPSS.

Online Study Resources

SociologyNow™: Research Methods

1. Before you do your final review of the chapter, take the *SociologyNow: Research Methods* diagnostic quiz to help identify the areas on which you should concentrate. You'll find information on this online tool, as well as instructions on how to access all of its great resources, in the front of the book.
2. As you review, take advantage of the *Sociology Now: Research Methods* customized study plan, based on your quiz results. Use this study plan with its interactive exercises and other resources to master the material.
3. When you're finished with your review, take the posttest to confirm that you're ready to move on to the next chapter.

WEBSITE FOR *THE PRACTICE OF SOCIAL RESEARCH* 11TH EDITION

Go to your book's website at http://sociology.wadsworth.com/babbie_practice11e for tools to aid you in studying for your exams. You'll find *Tutorial Quizzes* with feedback, *Internet Exercises*, *Flashcards*, and *Chapter Tutorials*, as well as *Extended Projects*, *InfoTrac College Edition* search terms, *Social Research in Cyberspace*, *GSS Data*, *Web Links*, and primers for using various data-analysis software such as SPSS and NVivo.

WEB LINKS FOR THIS CHAPTER

Please realize that the Internet is an evolving entity, subject to change. Nevertheless, these few websites should be fairly stable. Also, check your book's website for even more *Web Links*. These websites, current at the time of this book's publication, provide opportunities to learn about conceptualization, operationalization, and measurement.

US Census, Statistical Abstract of the United States
http://www.census.gov/prod/www/statistical-abstract-us.html
Here is just about everything you want to know about people in the United States: what they are like and what they do. It provides numerous examples of how characteristics and behaviors can be defined and measured.

University of Michigan, General Social Survey Codebook
http://www.icpsr.umich.edu/GSS/
This is a major social science resource. The GSS codebook identifies the numerous variables examined by the studies over time and gives the specific operationalization of those variables.

University of Colorado, Social Science Data Archives
http://socsci.colorado.edu/LAB/dataarchives.htm
These hotlinks to major social science data sets will give you many examples of variables defined and studied by researchers.

CHAPTER SIX

Indexes, Scales, and Typologies

CHAPTER OVERVIEW

Researchers often need to employ multiple indicators to measure a variable adequately and validly. Indexes, scales, and typologies are useful composite measures made up of several indicators of variables.

SociologyNow™: Research Methods

Use this online tool to help you make the grade on your next exam. After reading this chapter, go to the "Online Study Resources" at the end of the chapter for instructions on how to benefit from *SociologyNow: Research Methods*.

Introduction

As we saw in Chapter 5, many social scientific concepts have complex and varied meanings. Making measurements that capture such concepts can be a challenge. Recall our discussion of content validity, which concerns whether we have captured all the different dimensions of a concept.

To achieve broad coverage of the various dimensions of a concept, we usually need to make multiple observations pertaining to that concept. Thus, for example, Bruce Berg (1989: 21) advises in-depth interviewers to prepare essential questions, which are "geared toward eliciting specific, desired information." In addition, the researcher should prepare extra questions: "questions roughly equivalent to certain essential ones, but worded slightly differently."

Multiple indicators are used with quantitative data as well. Suppose you're designing a survey. Although you can sometimes construct a single questionnaire item that captures the variable of interest—"Gender: ☐ Male ☐ Female" is a simple example—other variables are less straightforward and may require you to use several questionnaire items to measure them adequately.

Quantitative data analysts have developed specific techniques for combining indicators into a single measure. This chapter discusses the construction of two types of composite measures of variables—indexes and scales. Although these measures can be used in any form of social research, they are most common in survey research and other quantitative methods. A short section at the end of this chapter considers typologies, which are relevant to both qualitative and quantitative research.

Composite measures are frequently used in quantitative research, for several reasons. First, social scientists often wish to study variables that have no clear and unambiguous single indicators. Single indicators do suffice for some variables, such as age. We can determine a survey respondent's age by simply asking, "How old are you?" Similarly, we can determine a newspaper's circulation by merely looking at the figure the newspaper reports. In the case of complex concepts, however, researchers can seldom develop single indicators before they actually do the research. This is especially true with regard to attitudes and orientations. Rarely can a survey researcher, for example, devise single questionnaire items that adequately tap respondents' degrees of prejudice, religiosity, political orientations, alienation, and the like. More likely, the researcher will devise several items, each of which provides some indication of the variables. Taken individually, each of these items is likely to prove invalid or unreliable for many respondents. A composite measure, however, can overcome this problem.

Second, researchers may wish to employ a rather refined ordinal measure of a particular variable (*alienation*, say), arranging cases in several ordinal categories from very low to very high, for example. A single data item might not have enough categories to provide the desired range of variation. However, an index or scale formed from several items can provide the needed range.

Finally, indexes and scales are efficient devices for data analysis. If considering a single data item gives us only a rough indication of a given variable, considering several data items can give us a more comprehensive and more accurate indication. For example, a single newspaper editorial may give us some indication of the political orientations of that newspaper. Examining several editorials would probably give us a better assessment, but the manipulation of several data items simultaneously could be very complicated. Indexes and scales (especially scales) are efficient data-reduction devices: They allow us to summarize several indicators in a single numerical score, while sometimes nearly maintaining the specific details of all the individual indicators.

Indexes versus Scales

The terms *index* and *scale* are typically used imprecisely and interchangeably in social research literature. The two types of measures do have some characteristics in common, but in this book we'll distinguish between the two. However, you should

be warned of a growing tendency in the literature to use the term *scale* to refer to both indexes and scales, as they are distinguished here.

First, let's consider what they have in common. Both scales and indexes are ordinal measures of variables. Both rank-order the units of analysis in terms of specific variables such as *religiosity, alienation, socioeconomic status, prejudice,* or *intellectual sophistication*. A person's score on either a scale or an index of religiosity, for example, gives an indication of his or her relative religiosity vis-à-vis other people.

Further, both scales and indexes are composite measures of variables—that is, measurements based on more than one data item. Thus, a survey respondent's score on an index or scale of religiosity is determined by the responses given to several questionnaire items, each of which provides some indication of religiosity. Similarly, a person's IQ score is based on answers to a large number of test questions. The political orientation of a newspaper might be represented by an index or scale score reflecting the newspaper's editorial policy on various political issues.

Despite these shared characteristics, it's useful to distinguish between indexes and scales. In this book, we'll distinguish them by the way scores are assigned in each. We construct an **index** simply by accumulating scores assigned to individual attributes. We might measure prejudice, for example, by adding up the number of prejudiced statements each respondent agreed with. We construct a **scale**, however, by assigning scores to patterns of responses, recognizing that some items reflect a relatively weak degree of the variable while others reflect something stronger. For example, agreeing that "Women are different from men" is, at best, weak evidence of sexism compared with agreeing that "Women should not be allowed to vote." A scale takes advantage of differences in intensity among the attributes of the same variable to identify distinct patterns of response.

Let's consider this simple example of sexism a bit further. Imagine asking people to agree or disagree with the two statements just presented. Some might agree with both, some might disagree with both. But suppose I told you someone agreed with one and disagree with the other: Could you guess which statement they agreed with and which they did not? I'd guess the person in question agreed that women were different but disagreed that they should be prohibited from voting. On the other hand, I doubt that anyone would want to prohibit women from voting, while asserting that there is no difference between men and women. That would make no sense.

Now consider this. The two responses we wanted from each person would technically yield four response patterns: agree/agree, agree/disagree, disagree/agree, and disagree/disagree. We've just seen, however, that only three of the four patterns make any sense or are likely to occur. Where indexes score people based on their *responses*, scales score people on the basis of *response patterns*: We determine what the logical response patterns are and score people in terms of the pattern their responses most closely resemble.

Figure 6-1 provides a graphic illustration of the difference between indexes and scales. Let's assume we want to develop a measure of political activism, distinguishing those people who are very active in political affairs, those who don't participate much at all, and those who are somewhere in between.

The first part of Figure 6-1 illustrates the logic of indexes. The figure shows six different political actions. Although you and I might disagree on some specifics, I think we could agree that the six actions represent roughly the same degree of political activism.

Using these six items, we could construct an index of political activism by giving each person 1 point for each of the actions he or she has taken. If you wrote to a public official and signed a petition,

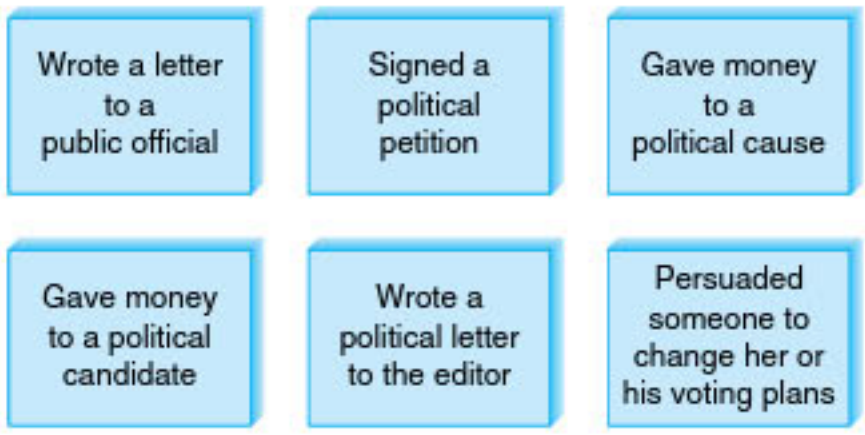
index A type of composite measure that summarizes and rank-orders several specific observations and represents some more general dimension.

scale A type of composite measure composed of several items that have a logical or empirical structure among them. Examples of scales include Bogardus social distance, Guttman, Likert, and Thurstone scales.

Index-Construction Logic

Here are several types of political actions people may have taken. By and large, the different actions represent similar *degrees* of political activism.

To create an *index* of overall political activism, we might give people 1 point for each of the actions they've taken.



Scale-Construction Logic

Here are some political actions that represent very different degrees of activism: e.g., running for office represents a higher degree of activism than simply voting does. It seems likely, moreover, that anyone who has taken one of the more demanding actions would have taken all the easier ones as well.

To construct a *scale* of political activism, we might score people according to which of the following "ideal" patterns comes closest to describing them.

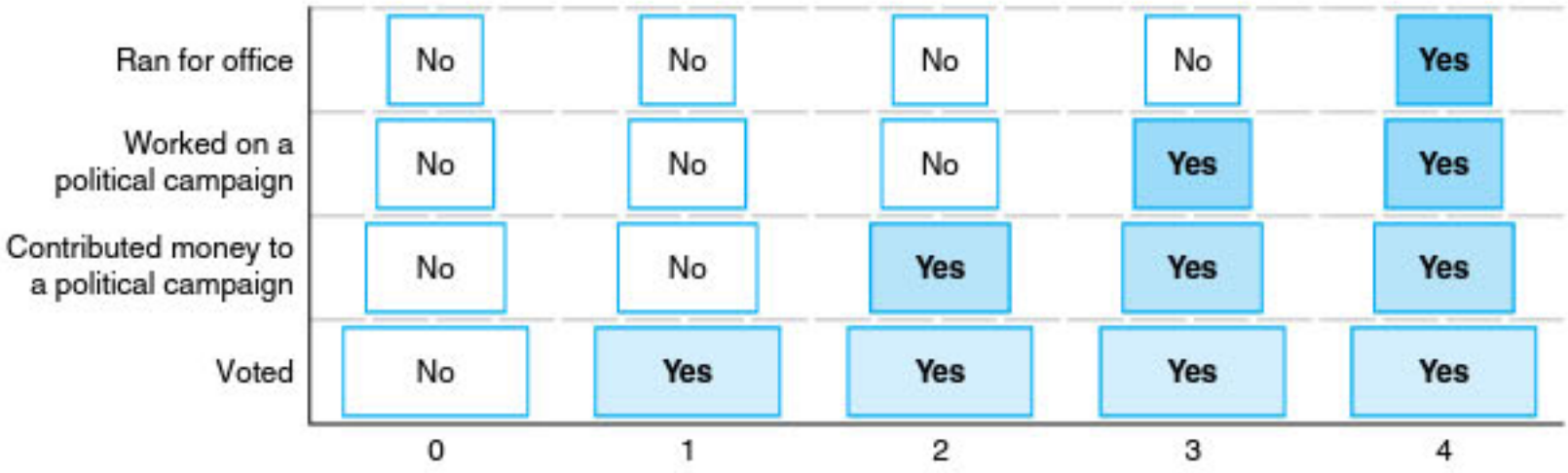


FIGURE 6-1

Indexes versus Scales. Both indexes and scales seek to measure variables such as political activism. Whereas indexes count the number of indicators of the variable, scales take account of the differing intensities of those indicators.

you'd get a total of 2 points. If I gave money to a candidate and persuaded someone to change her or his vote, I'd get the same score as you. Using this approach, we'd conclude that you and I had the same degree of political activism, even though we had taken different actions.

The second part of Figure 6-1 describes the logic of scale construction. In this case, the actions clearly represent different degrees of political activism, ranging from simply voting to running for office. Moreover, it seems safe to assume a pattern of actions in this case. For example, all those who contributed money probably also voted. Those who worked on a campaign probably also gave some money and voted. This suggests that most people will fall into only one of five idealized action patterns, represented by the illustrations at the bottom of the figure. The discussion of scales, later in this chapter, describes ways of identifying people with the type they most closely represent.

As you might surmise, scales are generally superior to indexes, because scales take into consideration the intensity with which different items reflect the variable being measured. Also, as the example in Figure 6-1 shows, scale scores convey more information than index scores do. Again,

be aware that the term *scale* is commonly misused to refer to measures that are only indexes. Merely calling a measure a scale instead of an index doesn't make it better.

There are two other misconceptions about scaling that you should know about. First, whether the combination of several data items results in a scale almost always depends on the particular sample of observations under study. Certain items may form a scale within one sample but not within another. For this reason, do not assume that a given set of items is a scale simply because it has turned out that way in an earlier study.

Second, the use of specific scaling techniques—such as Guttman scaling, to be discussed—does not ensure the creation of a scale. Rather, such techniques let us determine whether or not a set of items constitutes a scale.

An examination of actual social science research reports will show that researchers use indexes much more frequently than they do scales. Ironically, however, the methodological literature contains little if any discussion of index construction, whereas discussions of scale construction abound. There appear to be two reasons for this disparity. First, indexes are more frequently used because scales are often difficult or impossible to construct from the data at hand. Second, methods of index construction seem so obvious and straightforward that they aren't discussed much.

Constructing indexes is not a simple undertaking, however. The general failure to develop index construction techniques has resulted in many bad indexes in social research. With this in mind, I've devoted over half of this chapter to the methods of index construction. With a solid understanding of the logic of this activity, you'll be better equipped to try constructing scales. Indeed, a carefully constructed index may turn out to be a scale.

Index Construction

Let's look now at four main steps in the construction of an index: selecting possible items, examining their empirical relationships, scoring the index, and validating it. We'll conclude this discussion by examining the construction of an index that provided interesting findings about the status of women in different countries.

Item Selection

The first step in creating an index is selecting items for a composite index, which is created to measure some variable.

Face Validity

The first criterion for selecting items to be included in an index is face validity (or logical validity). If you want to measure political conservatism, for example, each of your items should appear on its face to indicate conservatism (or its opposite, liberalism). Political party affiliation would be one such item. Another would be an item asking people to approve or disapprove of the views of a well-known conservative public figure. In constructing an index of religiosity, you might consider items such as attendance at religious services, acceptance of certain religious beliefs, and frequency of prayer; each of these appears to offer some indication of religiosity.

Unidimensionality

The methodological literature on conceptualization and measurement stresses the need for unidimensionality in scale and index construction. That is, a composite measure should represent only one dimension of a concept. Thus, items reflecting religious fundamentalism should not be included in a measure of political conservatism, even though the two variables might be empirically related to each other.

General or Specific

Although measures should tap the same dimension, the general dimension you're attempting to measure may have many nuances. In the example of religiosity, the indicators mentioned previously—ritual participation, belief, and so on—represent different types of religiosity. If you wished to focus

on ritual participation in religion, you should choose items specifically indicating this type of religiosity: attendance at religious services and other rituals such as confession, bar mitzvah, bowing toward Mecca, and the like. If you wished to measure religiosity in a more general way, you would include a balanced set of items, representing each of the different types of religiosity. Ultimately, the nature of the items you include will determine how specifically or generally the variable is measured.

Variance

In selecting items for an index, you must also be concerned with the amount of variance they provide. If an item is intended to indicate political conservatism, for example, you should note what proportion of respondents would be identified as conservatives by that item. If a given item identified no one as a conservative or everyone as a conservative—for example, if nobody indicated approval of a radical-right political figure—that item would not be very useful in the construction of an index.

To guarantee variance, you have two options. First, you may select several items the responses to which divide people about equally in terms of the variable, for example, about half conservative and half liberal. Although no single response would justify the characterization of a person as very conservative, a person who responded as a conservative on all items might be so characterized.

The second option is to select items differing in variance. One item might identify about half the subjects as conservative, while another might identify few of the respondents as conservatives. Note that this second option is necessary for scaling, and it is reasonable for index construction as well.

Examination of Empirical Relationships

The second step in index construction is to examine the empirical relationships among the items being considered for inclusion. (See Chapter 14 for more.) An empirical relationship is established when respondents' answers to one question—in a questionnaire, for example—help us predict how they'll answer other questions. If two items are empirically related to each other, we can reasonably argue that each reflects the same variable, and we may include them both in the same index. There are two types of possible relationships among items: bivariate and multivariate.

Bivariate Relationships

A *bivariate relationship* is, simply put, a relationship between two variables. Suppose we want to measure respondents' support for U.S. participation in the United Nations. One indicator of different levels of support might be the question "Do you feel the U.S. financial support of the UN is ☐ Too high ☐ About right ☐ Too low?"

A second indicator of support for the United Nations might be the question "Should the United States contribute military personnel to UN peacekeeping actions? ☐ Strongly approve ☐ Mostly approve ☐ Mostly disapprove ☐ Strongly disapprove."

Both of these questions, on their face, seem to reflect different degrees of support for the United Nations. Nonetheless, some people might feel the United States should give more money but not provide troops. Others might favor sending troops but cutting back on financial support.

If the two items both reflect degrees of the same thing, however, we should expect responses to the two items to generally correspond with each other. Specifically, those who approve of military support should be more likely to favor financial support than those who disapprove of military support would. Conversely, those who favor financial support should be more likely to favor military support than those disapproving of financial support would. If these expectations are met, we say there is a bivariate relationship between the two items.

Here's another example. Suppose we want to determine the degree to which respondents feel women have the right to an abortion. We might ask (1) "Do you feel a woman should have the right to an abortion when her pregnancy was the result of rape?" and (2) "Do you feel a woman should have the right to an abortion if continuing her pregnancy would seriously threaten her life?"

"Cause" and "Effect" Indicators

by Kenneth Bollen

Department of Sociology, University of North Carolina, Chapel Hill

While it often makes sense to expect indicators of the same variable to be positively related to one another, as discussed in the text, this is not always the case.

Indicators should be related to one another if they are essentially "effects" of a variable. For example, to measure self-esteem, we might ask a person to indicate whether he or she agrees or disagrees with the statements (1) "I am a good person" and (2) "I am happy with myself." A person with high self-esteem should agree with both statements while one with low self-esteem would probably disagree with both. Since each indicator depends on or "reflects" self-esteem, we expect them to be positively correlated. More generally, indicators that depend on the same variable should be associated with one another if they are valid measures.

But, this is not the case when the indicators are the "cause" rather than the "effect" of a variable. In this situation the indicators may correlate positively, negatively, or not at all. For example, we could use gender and race as indicators of the variable *exposure to discrimination*. Being nonwhite or female increases the likelihood of experiencing discrimination, so both are good indicators of the variable. But we would not expect the race and gender of individuals to be strongly associated.

Or, we may measure *social interaction* with three indicators: time spent with friends, time spent with family, and time spent with coworkers. Though each indicator is valid, they need not be positively correlated. Time spent with friends, for instance, may be inversely related to time spent with family. Here, the three indicators "cause" the degree of social interaction.

As a final example, *exposure to stress* may be measured by whether a person recently experienced divorce, death of a spouse, or loss of a job. Though any of these events may indicate stress, they need not correlate with one another.

In short, we expect an association between indicators that depend on or "reflect" a variable, that is, if they are the "effects" of the variable. But if the variable depends on the indicators—if the indicators are the "causes"—those indicators may be either positively or negatively correlated, or even unrelated. Therefore, we should decide whether indicators are causes or effects of a variable before using their intercorrelations to assess their validity.

Granted, some respondents might agree with item (1) and disagree with item (2); others will do just the reverse. However, if both items tap into some general opinion people have about the issue of abortion, then the responses to these two items should be related to each other. Those who support the right to an abortion in the case of rape should be more likely to support it if the woman's life is threatened than those who disapproved of abortion in the case of rape would. This would be another example of a bivariate relationship.

You should examine all the possible bivariate relationships among the several items being considered for inclusion in an index, in order to determine the relative strengths of relationships among the several pairs of items. Percentage tables, correlation coefficients (see Chapter 16), or both may be used for this purpose. How we evaluate the strength of the relationships, however, can be rather subtle. "'Cause' and 'Effect' Indicators" examines some of these subtleties.

Be wary of items that are not related to one another empirically: It's unlikely that they measure the same variable. You should probably drop any item that is not related to several other items.

At the same time, a very strong relationship between two items presents a different problem. If two items are perfectly related to each other, then only one needs to be included in the index; because it completely conveys the indications provided by the other, nothing more would be added by including the other item. (This problem will become even clearer in the next section.)

Here's an example to illustrate the testing of bivariate relationships in index construction. I once conducted a survey of medical school faculty members to find out about the consequences of a "scientific perspective" on the quality of patient care

provided by physicians. The primary intent was to determine whether scientifically inclined doctors treated patients more impersonally than other doctors did.

The survey questionnaire offered several possible indicators of respondents' scientific perspectives. Of those, three items appeared to provide especially clear indications of whether the doctors were scientifically oriented:

1. As a medical school faculty member, in what capacity do you feel you can make your greatest *teaching* contribution: as a practicing physician or as a medical researcher?
2. As you continue to advance your own medical knowledge, would you say your ultimate medical interests lie primarily in the direction of total patient management or the understanding of basic mechanisms? [The purpose of this item was to distinguish those who were mostly interested in overall patient care from those mostly interested in biological processes.]
3. In the field of therapeutic research, are you generally more interested in articles reporting evaluations of the effectiveness of various treatments or articles exploring the basic rationale underlying the treatments? [Similarly, I wanted to distinguish those more interested in articles dealing with patient care from those more interested in biological processes.]

(Babbie 1970: 27–31)

For each of these items, we might conclude that those respondents who chose the second answer are more scientifically oriented than respondents who chose the first answer. Though this comparative conclusion is reasonable, we should not be misled into thinking that respondents who chose the second answer to a given item are scientists in any absolute sense. They are simply more scientifically oriented than those who chose the first answer to the item.

To see this point more clearly, let's examine the distribution of responses to each item. From the first item—greatest teaching contribution—only about one-third of the respondents appeared scientifically oriented. That is, approximately one-third said they could make their greatest teaching contribution as medical researchers. In response to the second item—ultimate medical interests—approximately two-thirds chose the scientific answer, saying they were more interested in learning about basic mechanisms than learning about total patient management. In response to the third item—reading preferences—about 80 percent chose the scientific answer.

These three questionnaire items can't tell us how many "scientists" there are in the sample, for none of them is related to a set of criteria for what constitutes being a scientist in any absolute sense. Using the items for this purpose would present us with the problem of three quite different estimates of how many scientists there were in the sample.

However, these items do provide us with three independent indicators of respondents' relative inclinations toward science in medicine. Each item separates respondents into the more scientific and the less scientific. But each grouping of more or less scientific respondents will have a somewhat different membership from the others. Respondents who seem scientific in terms of one item will not seem scientific in terms of another. Nevertheless, to the extent that each item measures the same general dimension, we should find some correspondence among the several groupings. Respondents who appear scientific in terms of one item should be more likely to appear scientific in their response to another item than would those who appeared nonscientific in their response to the first. In other words, we should find an association or correlation between the responses given to two items.

Figure 6-2 shows the associations among the responses to the three items. Three bivariate tables are presented, showing the distribution of responses for each possible pairing of items. An examination of the three bivariate relationships presented in the figure supports the suggestion that the three items all measure the same variable: *scientific orientation*. To see why this is so, let's begin by looking at the first bivariate relationship in the table. The table shows that faculty who responded that "researcher" was the role in which they could make their greatest teaching contribution were more likely to identify their ultimate medical interests as "basic mechanisms" (87 percent) than

a.

Greatest Teaching Contribution

Ultimate Medical Interest	Physician	Researcher
Total patient management	49%	13%
Basic mechanisms	51%	87%
	100% (268)	100% (159)

b.

Reading Preferences

Ultimate Medical Interest	Effectiveness	Rationale
Total patient management	68%	30%
Basic mechanisms	32%	70%
	100% (78)	100% (349)

c.

Reading Preferences

Greatest Teaching Contribution	Effectiveness	Rationale
Physician	85%	64%
Researcher	15%	36%
	100% (78)	100% (349)

FIGURE 6-2

Bivariate Relationships among Scientific Orientation Items. If several indicators are measures of the same variable, then they should be empirically correlated with one another.

were those who answered "physician" (51 percent). The fact that the "physicians" are about evenly split in their ultimate medical interests is irrelevant for our purposes. It is only relevant that they are less scientific in their medical interests than the "researchers." The strength of this relationship may be summarized as a 36 percentage point difference.

The same general conclusion applies to the other bivariate relationships. The strength of the relationship between reading preferences and ultimate medical interests may be summarized as a 38 percentage point difference, and the strength of the relationship between reading preferences and the two teaching roles as a 21 percentage point difference. In summary, then, each single item produces a different grouping of "scientific" and "nonscientific" respondents. However, the responses given to each of the items correspond, to a greater or lesser degree, to the responses given to each of the other items.

Initially, the three items were selected on the basis of face validity—each appeared to give some indication of faculty members' orientations to science. By examining the bivariate relationship between the pairs of items, we have found support for the expectation that they all measure basically the same thing. However, that support does not sufficiently justify including the items in a composite index. Before combining them in a single index, we need to examine the multivariate relationships among the several variables.

Multivariate Relationships among Items

Figure 6-3 categorizes the sample respondents into four groups according to (1) their greatest teaching contribution and (2) their reading preferences. The numbers in parentheses indicate the number of respondents in each group. Thus, 66 of the faculty members who said they could best teach as physicians also said they preferred articles dealing with the effectiveness of treatments. For each of the four groups, the figure presents the percentage of those who say they are ultimately more interested in basic mechanisms. So, for example, of the 66 faculty mentioned, 27 percent are primarily interested in basic mechanisms.

The arrangement of the four groups is based on a previously drawn conclusion regarding scientific orientations. The group in the upper left corner of the table is presumably the least scientifically oriented, based on greatest teaching contribution and reading preference. The group in the lower right corner is presumably the most scientifically oriented in terms of those items.

Percent Interested in Basic Mechanisms

Reading Preferences	Greatest Teaching Contribution: Physician	Greatest Teaching Contribution: Researcher
Effectiveness of treatments	27% (66)	58% (12)
Rationale behind treatments	58% (219)	89% (130)

FIGURE 6-3

Trivariate Relationships among Scientific Orientation Items. Indicators of the same variable should be correlated in a multivariate analysis as well as in bivariate analyses.

Percent Interested in Basic Mechanisms

Reading Preferences	Greatest Teaching Contribution: Physician	Greatest Teaching Contribution: Researcher
Effectiveness of treatments	51% (66)	87% (12)
Rationale behind treatments	51% (219)	87% (130)

FIGURE 6-4

Hypothetical Trivariate Relationship among Scientific Orientation Items. This hypothetical relationship would suggest that not all three indicators would contribute effectively to a composite index.

Recall that expressing a primary interest in basic mechanisms was also taken as an indication of scientific orientation. As we should expect, then, those in the lower right corner are the most likely to give this response (89 percent), and those in the upper left corner are the least likely (27 percent). The respondents who gave mixed responses in terms of teaching contributions and reading preferences have an intermediate rank in their concern for basic mechanisms (58 percent in both cases).

This table tells us many things. First, we may note that the original relationships between pairs of items are not significantly affected by the presence of a third item. Recall, for example, that the relationship between teaching contribution and ultimate medical interest was summarized as a 36 percentage point difference. Looking at Figure 6-3, we see that among only those respondents who are most interested in articles dealing with the effectiveness of treatments, the relationship between teaching contribution and ultimate medical interest is 31 percentage points (58 percent minus 27 percent: first row). The same is true among those most interested in articles dealing with the rationale for treatments (89 percent minus 58 percent: second row). The original relationship between teaching contribution and ultimate medical interest is essentially the same as in Figure 6-2, even among those respondents judged as scientific or nonscientific in terms of reading preferences.

We can draw the same conclusion from the columns in Figure 6-3. Recall that the original relationship between reading preferences and ultimate medical interests was summarized as a 38 percentage point difference. Looking only at the "physicians" in Figure 6-3, we see that the relationship between the other two items is now 31 percentage points. The same relationship is found among the "researchers" in the second column.

The importance of these observations becomes clearer when we consider what might have happened. In Figure 6-4, hypothetical data tell a much different story than the actual data in Figure 6-3 do. As you can see, Figure 6-4 shows that the original relationship between teaching role and ultimate medical interest persists, even when reading preferences are introduced into the picture. In each row of the table, the "researchers" are more likely to express an interest in basic mechanisms than the "physicians" are. Looking down the columns, however, we note that there is no relationship between reading preferences and ultimate medical interest. If we know whether a respondent feels he or she can best teach as a physician or as a researcher, knowing the respondent's reading preference adds nothing to our evaluation of his or her scientific orientation. If something like Figure 6-4 resulted from the actual data, we would conclude that reading preference should not be included in the same index as teaching role, because it contributed nothing to the composite index.

This example used only three questionnaire items. If more were being considered, then more-complex multivariate tables would be in order, constructed of four, five, or more variables. The purpose of this step in index construction, again, is to discover the simultaneous interaction of the items in order to determine which should be included in the same index. These kinds of data analyses are easily accomplished using programs such as SPSS and MicroCase. They are usually referred to as cross-tabulations.

Index Scoring

When you've chosen the best items for your index, you next assign scores for particular responses, thereby creating a single composite measure out of the several items. There are two basic decisions to be made in this step.

First, you must decide the desirable range of the index scores. A primary advantage of an index over a single item is the range of gradations it offers in the measurement of a variable. As noted earlier, political conservatism might be measured from "very conservative" to "not at all conservative" or "very liberal." How far to the extremes, then, should the index extend?

In this decision, the question of variance enters once more. Almost always, as the possible extremes of an index are extended, fewer cases are to be found at each end. The researcher who wishes to measure political conservatism to its greatest extreme (somewhere to the right of Attila the Hun, as the saying goes) may find there is almost no one in that category. At some point, additional gradations do not add meaning to the results.

The first decision, then, concerns the conflicting desire for (1) a range of measurement in the index and (2) an adequate number of cases at each point in the index. You'll be forced to reach some kind of compromise between these conflicting desires.

The second decision concerns the actual assignment of scores for each particular response. Basically you must decide whether to give items in the index equal weight or different weights. Although there are no firm rules, I suggest—and practice tends to support this method—that items be weighted equally unless there are compelling reasons for differential weighting. That is, the burden of proof should be on differential weighting; equal weighting should be the norm.

Of course, this decision must be related to the earlier issue regarding the balance of items chosen. If the index is to represent the composite of slightly different aspects of a given variable, then you should give each aspect the same weight. In some instances, however, you may feel that two items reflect essentially the same aspect, and the third reflects a different aspect. If you want to have both aspects equally represented by the index, you might give the different item a weight equal to the combination of the two similar ones. For instance, you could assign a maximum score of 2 to the different item and a maximum score of 1 to each of the similar ones.

Although the rationale for scoring responses should take such concerns as these into account, typically researchers experiment with different scoring methods, examining the relative weights given to different aspects but at the same time worrying about the range and distribution of cases provided. Ultimately, the scoring method chosen will represent a compromise among these several demands. Of course, as in most research activities, such a decision is open to revision on the basis of later examinations. Validation of the index, to be discussed shortly, may lead the researcher to recycle his or her efforts toward constructing a completely different index.

In the example taken from the medical school faculty survey, I decided to weight the items equally, because I'd chosen them, in part, because they represent slightly different aspects of the overall variable *scientific orientation*. On each of the items, the respondents were given a score of 1 for choosing the "scientific" response to the item and a score of 0 for choosing the "nonscientific" response. Each respondent, then, could receive a score of 0, 1, 2, or 3. This scoring method provided what I considered a useful range of variation—four index categories—and also provided enough cases for analysis in each category.

Here's a similar example of index scoring, from a study of work satisfaction. One of the key

variables was *job-related depression*, measured by an index composed of the following four items, which asked workers how they felt when thinking about themselves and their jobs:

- "I feel downhearted and blue."
- "I get tired for no reason."
- "I find myself restless and can't keep still."
- "I am more irritable than usual."

The researchers, Amy Wharton and James Baron, report, "Each of these items was coded: 4 = often, 3 = sometimes, 2 = rarely, 1 = never." They go on to explain how they measured another variable, *job-related self-esteem*:

> Job-related self-esteem was based on four items asking respondents how they saw themselves in their work: happy/sad; successful/not successful; important/not important; doing their best/not doing their best. Each item ranged from 1 to 7, where 1 indicates a self-perception of not being happy, successful, important, or doing one's best.
>
> *(1987: 578)*

As you look through the social research literature, you'll find numerous similar examples of cumulative indexes being used to measure variables. Sometimes the indexing procedures are controversial, as evidenced in "What Is the Best College in the United States?"

Handling Missing Data

Regardless of your data-collection method, you'll frequently face the problem of missing data. In a content analysis of the political orientations of newspapers, for example, you may discover that a particular newspaper has never taken an editorial position on one of the issues being studied. In an experimental design involving several retests of subjects over time, some subjects may be unable to participate in some of the sessions. In virtually every survey, some respondents fail to answer some questions (or choose a "don't know" response). Although missing data present problems at all stages of analysis, they're especially troublesome in index construction. There are, however, several methods of dealing with these problems.

First, if there are relatively few cases with missing data, you may decide to exclude them from the construction of the index and the analysis. (I did this in the medical school faculty example.) The primary concerns in this instance are whether the numbers available for analysis will remain sufficient and whether the exclusion will result in an unrepresentative sample whenever the index, excluding some of the respondents, is used in the analysis. The latter possibility can be examined through a comparison—on other relevant variables—of those who would be included and excluded from the index.

Second, you may sometimes have grounds for treating missing data as one of the available responses. For example, if a questionnaire has asked respondents to indicate their participation in various activities by checking "yes" or "no" for each, many respondents may have checked some of the activities "yes" and left the remainder blank. In such a case, you might decide that a failure to answer meant "no," and score missing data in this case as though the respondents had checked the "no" space.

Third, a careful analysis of missing data may yield an interpretation of their meaning. In constructing a measure of political conservatism, for example, you may discover that respondents who failed to answer a given question were generally as conservative on other items as those who gave the conservative answer. In another example, a recent study measuring religious beliefs found that people who answered "don't know" about a given belief were almost identical to the "disbelievers" in their answers about other beliefs. (*Note*: You should take these examples not as empirical guides in your own studies but only as suggestions of general ways to analyze your own data.) Whenever the analysis of missing data yields such interpretations, then, you may decide to score such cases accordingly.

There are many other ways of handling the problem of missing data. If an item has several possible values, you might assign the middle value to cases with missing data; for example, you could assign a 2 if the values are 0, 1, 2, 3, and 4. For a

What Is the Best College in the United States?

Each year the newsmagazine *U.S. News and World Report* issues a special report ranking the nation's colleges and universities. Their rankings reflect an index, created from several items: educational expenditures per student, graduation rates, selectivity (percentage accepted of those applying), average SAT scores of first-year students, and similar indicators of quality.

Typically, Harvard is ranked the number one school in the nation, followed by Yale and Princeton. However, the 1999 "America's Best Colleges" issue shocked educators, prospective college students, and their parents. The California Institute of Technology had leaped from ninth place in 1998 to first place a year later. While Harvard, Yale, and Princeton still did well, they had been supplanted. What had happened at Caltech to produce such a remarkable surge in quality?

The answer was to be found at *U.S. News and World Report*, not at Caltech. The newsmagazine changed the structure of the ranking index in 1999, which made a big difference in how schools fared.

Bruce Gottlieb (1999) gives this example of how the altered scoring made a difference.

> So, how did Caltech come out on top? Well, one variable in a school's ranking has long been educational expenditures per student, and Caltech has traditionally been tops in this category. But until this year, *U.S. News* considered only a school's ranking in this category—first, second, etc.—rather than how much it spent relative to other schools. It didn't matter whether Caltech beat Harvard by $1 or by $100,000. Two other schools that rose in their rankings this year were MIT (from fourth to third) and Johns Hopkins (from 14th to seventh). All three have high per-student expenditures and all three are especially strong in the hard sciences. Universities are allowed to count their research budgets in their per-student expenditures, though students get no direct benefit from costly research their professors are doing outside of class.
>
> In its "best colleges" issue two years ago, *U.S. News* made precisely this point, saying it considered only the rank ordering of per-student expenditures, rather than the actual amounts, on the grounds that expenditures at institutions with large research programs and medical schools are substantially higher than those at the rest of the schools in the category. In other words, just two years ago, the magazine felt it unfair to give Caltech, MIT, and Johns Hopkins credit for having lots of fancy laboratories that don't actually improve undergraduate education.

Gottlieb reviewed each of the changes in the index and then asked how 1998's ninth-ranked Caltech would have done had the revised indexing formula been in place a year earlier. His conclusion: Caltech would have been first in 1998 as well. In other words, the apparent improvement was solely a function of how the index was scored.

Composite measures such as scales and indexes are valuable tools for understanding society. However, it's important that we know how those measures are constructed and what that construction implies.

So, what's *really* the best college in the United States? It depends on how you define "best." There is no "really best," only the various social constructions we can create.

Sources: U.S. News and World Report, "America's Best Colleges," August 30, 1999; Bruce Gottlieb, "Cooking the School Books: How *U.S. News* Cheats in Picking Its 'Best American Colleges,'" Slate, August 31, 1999 (http://slate.msn.com/default.aspx?id=34027).

continuous variable such as age, you could similarly assign the mean to cases with missing data (more on this in Chapter 14). Or, missing data can be supplied by assigning values at random. All of these are conservative solutions because they weaken the "purity" of your index and reduce the likelihood that it will relate to other variables in ways you may have hypothesized.

If you're creating an index out of a large number of items, you can sometimes handle missing data by using proportions based on what is observed. Suppose your index is composed of six indicators, and you only have four observations for a particular subject. If the subject has earned 4 points out of a possible 4, you might assign an index score of 6; if the subject has 2 points (half the possible score on four items), you could assign a score of 3 (half the possible score on six observations).

The choice of a particular method to be used depends so much on the research situation that I can't reasonably suggest a single "best" method or rank the several I've described. Excluding all cases with missing data can bias the representativeness of the findings, but including such cases by assigning

scores to missing data can influence the nature of the findings. The safest and best method is to construct the index using more than one of these methods and see whether you reach the same conclusions using each of the indexes. Understanding your data is the final goal of analysis anyway.

Index Validation

Up to this point, we've discussed all the steps in the selection and scoring of items that result in an index purporting to measure some variable. If each of the preceding steps is carried out carefully, the likelihood of the index actually measuring the variable is enhanced. To demonstrate success, however, we must show that the index is valid. Following the basic logic of validation, we assume that the index provides a measure of some variable; that is, the scores on the index arrange cases in a rank order in terms of that variable. An index of political conservatism rank-orders people in terms of their relative conservatism. If the index does that successfully, then people scored as relatively conservative on the index should appear relatively conservative in all other indications of political orientation, such as their responses to other questionnaire items. There are several methods of validating an index.

Item Analysis

The first step in index validation is an internal validation called **item analysis**. In item analysis, you examine the extent to which the index is related to (or predicts responses to) the individual items it comprises. Here's an illustration of this step.

In the index of scientific orientations among medical school faculty, index scores ranged from 0 (most interested in patient care) to 3 (most interested in research). Now let's consider one of the items in the index: whether respondents wanted to advance their own knowledge more with regard to total patient management or more in the area of basic mechanisms. The latter were treated as being more scientifically oriented than the former. The following empty table shows how we would examine the relationship between the index and the individual item.

	Index of Scientific Orientations			
	0	*1*	*2*	*3*
Percentage who said they were more interested in basic mechanisms	??	??	??	??

If you take a minute to reflect on the table, you may see that we already know the numbers that go in two of the cells. To get a score of 3 on the index, respondents had to say "basic mechanisms" in response to this question and give the "scientific" answers to the other two items as well. Thus, 100 percent of the 3's on the index said "basic mechanisms." By the same token, all the 0's had to answer this item with "total patient management." Thus, 0 percent of those respondents said "basic mechanisms." Here's how the table looks with the information we already know.

	Index of Scientific Orientations			
	0	*1*	*2*	*3*
Percentage who said they were more interested in basic mechanisms	0	??	??	100

If the individual item is a good reflection of the overall index, we should expect the 1's and 2's to fill in a progression between 0 percent and 100 percent. More of the 2's should choose "basic mechanisms" than 1's. This result is not guaranteed by the way the index was constructed, however; it is an empirical question—one we answer in an item analysis. Here's how this particular item analysis turned out.

	Index of Scientific Orientations			
	0	*1*	*2*	*3*
Percentage who said they were more interested in basic mechanisms	0	16	91	100

item analysis An assessment of whether each of the items included in a composite measure makes an independent contribution or merely duplicates the contribution of other items in the measure.

As you can see, in accord with our assumption that the 2's are more scientifically oriented than the 1's, we find that a higher percentage of the 2's (91 percent) say "basic mechanisms" than the 1's (16 percent).

An item analysis of the other two components of the index yields similar results, as shown here.

	Index of Scientific Orientations			
	0	*1*	*2*	*3*
Percentage who said they could teach best as medical researchers	0	4	14	100
Percentage who said they preferred reading about rationales	0	80	97	100

Each of the items, then, seems an appropriate component in the index. Each seems to reflect the same quality that the index as a whole measures.

In a complex index containing many items, this step provides a convenient test of the independent contribution of each item to the index. If a given item is found to be poorly related to the index, it may be assumed that other items in the index cancel out the contribution of that item, and it should be excluded from the index. If the item in question contributes nothing to the index's power, it should be excluded.

Although item analysis is an important first test of an index's validity, it is not a sufficient test. If the index adequately measures a given variable, it should successfully predict other indications of that variable. To test this, we must turn to items not included in the index.

External Validation

People scored as politically conservative on an index should appear conservative by other measures as well, such as their responses to other items in a questionnaire. Of course, we're talking about relative conservatism, because we can't define *conservatism* in any absolute way. However, those respondents scored as the most conservative on the index should score as the most conservative in answering other questions. Those scored as the least conservative on the index should score as the least conservative on other items. Indeed, the ranking of groups of respondents on the index should predict the ranking of those groups in answering other questions dealing with political orientations.

TABLE 6-1

Validation of Scientific Orientation Index

	Index of Scientific Orientation			
	Low *0*	*1*	*2*	*High* *3*
Percent interested in attending scientific lectures at the medical school	34	42	46	65
Percent who say faculty members should have experience as medical researchers	43	60	65	89
Percent who would prefer faculty duties involving research activities only	0	8	32	66
Percent who engaged in research during the preceding academic year	61	76	94	99

In our example of the scientific orientation index, several questions in the questionnaire offered the possibility of such **external validation**. Table 6-1 presents some of these items, which provide several lessons regarding index validation. First, we note that the index strongly predicts the responses to the validating items in the sense that the rank order of scientific responses among the four groups is the same as the rank order provided by the index itself. That is, the percentages reflect greater scientific orientation as you read across the rows of the table. At the same time, each item gives a different description of scientific orientations overall. For example, the last validating item indicates that the great majority of all faculty were engaged in

external validation The process of testing the validity of a measure, such as an index or scale, by examining its relationship to other, presumed indicators of the same variable. If the index really measures prejudice, for example, it should correlate with other indicators of prejudice.

research during the preceding year. If this were the only indicator of scientific orientation, we would conclude that nearly all faculty were scientific. Nevertheless, those scored as more scientific on the index are more likely to have engaged in research than were those who were scored as relatively less scientific. The third validating item provides a different descriptive picture: Only a minority of the faculty overall say they would prefer duties limited exclusively to research. Nevertheless, the relative percentages giving this answer correspond to the scores assigned on the index.

Bad Index versus Bad Validators

Nearly every index constructor at some time must face the apparent failure of external items to validate the index. If the internal item analysis shows inconsistent relationships between the items included in the index and the index itself, something is wrong with the index. But if the index fails to predict strongly the external validation items, the conclusion to be drawn is more ambiguous. In this situation we must choose between two possibilities: (1) the index does not adequately measure the variable in question, or (2) the validation items do not adequately measure the variable and thereby do not provide a sufficient test of the index.

Having worked long and conscientiously on the construction of an index, you'll likely find the second conclusion compelling. Typically, you'll feel you have included the best indicators of the variable in the index; the validating items are, therefore, second-rate indicators. Nevertheless, you should recognize that the index is purportedly a very powerful measure of the variable; thus, it should be somewhat related to any item that taps the variable even poorly.

When external validation fails, you should reexamine the index before deciding that the validating items are insufficient. One way to do this is to examine the relationships between the validating items and the individual items included in the index. If you discover that some of the index items relate to the validators and others do not, you'll have improved your understanding of the index as it was initially constituted.

There's no cookbook solution to this problem; it is an agony serious researchers must learn to survive. Ultimately, the wisdom of your decision to accept an index will be determined by the usefulness of that index in your later analyses. Perhaps you'll initially decide that the index is a good one and that the validators are defective, but you'll later find that the variable in question (as measured by the index) is not related to other variables in the ways you expected. You may then have to compose a new index.

The Status of Women: An Illustration of Index Construction

For the most part, our discussion of index construction has focused on the specific context of survey research, but other types of research also lend themselves to this kind of composite measure. For example, when the United Nations (1995) set out to examine the status of women in the world, they chose to create two indexes, reflecting two different dimensions.

The Gender-related Development Index (GDI) compared women to men in terms of three indicators: life expectancy, education, and income. These indicators are commonly used in monitoring the status of women in the world. The Scandinavian countries of Norway, Sweden, Finland, and Denmark ranked highest on this measure.

The second index, the Gender Empowerment Measure (GEM), aimed more at power issues and comprised three different indicators:

- The proportion of parliamentary seats held by women
- The proportion of administrative, managerial, professional, and technical positions held by women
- A measure of access to jobs and wages

Once again, the Scandinavian countries ranked high but were joined by Canada, New Zealand, the Netherlands, the United States, and Austria. Having two different measures of gender equality rather than one allowed the researchers to make more sophisticated distinctions. For example, in several

countries, most notably Greece, France, and Japan, women fared relatively well on the GDI but quite poorly on the GEM. Thus, while women were doing fairly well in terms of income, education, and life expectancy, they were still denied access to power. And whereas the GDI scores were higher in the wealthier nations than in the poorer ones, GEM scores showed that women's empowerment depended less on national wealth, with many poor, developing countries outpacing some rich, industrial ones in regard to such empowerment.

By examining several different dimensions of the variables involved in their study, the UN researchers also uncovered an aspect of women's earnings that generally goes unnoticed. Population Communications International (1996: 1) has summarized the finding nicely:

> Every year, women make an invisible contribution of eleven trillion U.S. dollars to the global economy, the UNDP [United Nations Development Programme] report says, counting both unpaid work and the underpayment of women's work at prevailing market prices. This "underevaluation" of women's work not only undermines their purchasing power, says the 1995 HDR [Human Development Report], but also reduces their already low social status and affects their ability to own property and use credit. Mahbub ul Haq, the principal author of the report, says that "if women's work were accurately reflected in national statistics, it would shatter the myth that men are the main breadwinner of the world." The UNDP report finds that women work longer hours than men in almost every country, including both paid and unpaid duties. In developing countries, women do approximately 53% of all work and spend two-thirds of their work time on unremunerated activities. In industrialized countries, women do an average of 51% of the total work, and—like their counterparts in the developing world—perform about two-thirds of their total labor without pay. Men in industrialized countries are compensated for two-thirds of their work.

As you can see, indexes can be constructed from many different kinds of data for a variety of purposes. Now we'll turn our attention from the construction of indexes to an examination of scaling techniques.

Scale Construction

Good indexes provide an ordinal ranking of cases on a given variable. All indexes are based on this kind of assumption: A senator who voted for seven conservative bills is considered to be more conservative than one who only voted for four of them. What an index may fail to take into account, however, is that not all indicators of a variable are equally important or equally strong. The first senator might have voted in favor of seven mildly conservative bills, whereas the second senator might have voted in favor of four extremely conservative bills. (The second senator might have considered the other seven bills too liberal and voted against them.)

Scales offer more assurance of ordinality by tapping the intensity structures among the indicators. The several items going into a composite measure may have different intensities in terms of the variable. Many methods of scaling are available. We'll look at four scaling procedures to illustrate the variety of techniques available, along with a technique called the semantic differential. Although these examples focus on questionnaires, the logic of scaling, like that of indexing, applies to other research methods as well.

Bogardus Social Distance Scale

Let's suppose you're interested in the extent to which U.S. citizens are willing to associate with, say, sex offenders. You might ask the following questions:

1. Are you willing to permit sex offenders to live in your country?
2. Are you willing to permit sex offenders to live in your community?
3. Are you willing to permit sex offenders to live in your neighborhood?

4. Would you be willing to let a sex offender live next door to you?
5. Would you let your child marry a sex offender?

These questions increase in terms of the closeness of contact with sex offenders. Beginning with the original concern to measure willingness to associate with sex offenders, you have thus developed several questions indicating differing degrees of intensity on this variable. The kinds of items presented constitute a **Bogardus social distance scale** (created by Emory Bogardus). This scale is a measurement technique for determining the willingness of people to participate in social relations—of varying degrees of closeness—with other kinds of people.

The clear differences of intensity suggest a structure among the items. Presumably if a person is willing to accept a given kind of association, he or she would be willing to accept all those preceding it in the list—those with lesser intensities. For example, the person who is willing to permit sex offenders to live in the neighborhood will surely accept them in the community and the nation but may or may not be willing to accept them as next-door neighbors or relatives. This, then, is the logical structure of intensity inherent among the items.

Empirically, one would expect to find the largest number of people accepting co-citizenship and the fewest accepting intermarriage. In this sense, we speak of "easy items" (for example, residence in the United States) and "hard items" (for example, intermarriage). More people agree to the easy items than to the hard ones. With some inevitable exceptions, logic demands that once a person has refused a relationship presented in the scale, he or she will also refuse all the harder ones that follow it.

The Bogardus social distance scale illustrates the important economy of scaling as a data-reduction device. By knowing how many relationships with sex offenders a given respondent will accept, we know which relationships were accepted. Thus, a single number can accurately summarize five or six data items without a loss of information.

Motoko Lee, Stephen Sapp, and Melvin Ray (1996) noticed an implicit element in the Bogardus social distance scale: It looks at social distance from the point of view of the majority group in a society. These researchers decided to turn the tables and create a "reverse social distance" scale: looking at social distance from the perspective of the minority group. Here's how they framed their questions (1996: 19):

> Considering typical Caucasian Americans you have known, not any specific person nor the worst or the best, circle Y or N to express your opinion.
>
> Y N 5. Do they mind your being a citizen in this country?
>
> Y N 4. Do they mind your living in the same neighborhood?
>
> Y N 3. Would they mind your living next to them?
>
> Y N 2. Would they mind your becoming a close friend to them?
>
> Y N 1. Would they mind your becoming their kin by marriage?

As with the original scale, the researchers found that knowing the number of items minority respondents agreed with also told the researchers which ones were agreed with: 98.9% percent of the time in this case.

Thurstone Scales

Often, the inherent structure of the Bogardus social distance scale is not appropriate to the variable being measured. Indeed, such a logical structure among several indicators is seldom apparent.

Bogardus social distance scale A measurement technique for determining the willingness of people to participate in social relations—of varying degrees of closeness—with other kinds of people. It is an especially efficient technique in that one can summarize several discrete answers without losing any of the original details of the data.

A **Thurstone scale** (created by Louis Thurstone) is an attempt to develop a format for generating groups of indicators of a variable that have at least an empirical structure among them. A group of judges is given perhaps a hundred items that are thought to be indicators of a given variable. Each judge is then asked to estimate how strong an indicator of a variable each item is—by assigning scores of perhaps 1 to 13. If the variable were *prejudice*, for example, the judges would be asked to assign the score of 1 to the very weakest indicators of prejudice, the score of 13 to the strongest indicators, and intermediate scores to those felt to be somewhere in between.

Once the judges have completed this task, the researcher examines the scores assigned to each item by all the judges, then determines which items produced the greatest agreement among the judges. Those items on which the judges disagreed broadly would be rejected as ambiguous. Among those items producing general agreement in scoring, one or more would be selected to represent each scale score from 1 to 13.

The items selected in this manner might then be included in a survey questionnaire. Respondents who appeared prejudiced on those items representing a strength of 5 would then be expected to appear prejudiced on those having lesser strengths, and if some of those respondents did not appear prejudiced on the items with a strength of 6, it would be expected that they would also not appear prejudiced on those with greater strengths.

If the Thurstone scale items were adequately developed and scored, the economy and effectiveness of data reduction inherent in the Bogardus social distance scale would appear. A single score might be assigned to each respondent (the strength of the hardest item accepted), and that score would adequately represent the responses to several questionnaire items. And as is true of the Bogardus scale, a respondent who scored 6 might be regarded as more prejudiced than one who scored 5 or less.

Thurstone scaling is not often used in research today, primarily because of the tremendous expenditure of energy and time required to have 10 to 15 judges score the items. Because the quality of their judgments would depend on their experience with the variable under consideration, they might need to be professional researchers. Moreover, the meanings conveyed by the several items indicating a given variable tend to change over time. Thus, an item having a given weight at one time might have quite a different weight later on. For a Thurstone scale to be effective, it would have to be periodically updated.

Thurstone scale A type of composite measure, constructed in accord with the weights assigned by "judges" to various indicators of some variables.

Likert Scaling

You may sometimes hear people refer to a questionnaire item containing response categories such as "strongly agree," "agree," "disagree," and "strongly disagree" as a *Likert scale*. This is technically a misnomer, although Rensis Likert (pronounced "LICK-ert") did create this commonly used question format.

The particular value of this format is the unambiguous ordinality of response categories. If respondents were permitted to volunteer or select such answers as "sort of agree," "pretty much agree," "really agree," and so forth, the researcher would find it impossible to judge the relative strength of agreement intended by the various respondents. The Likert format solves this problem.

Likert had something more in mind, however. He created a method by which this question format could be used to determine the relative intensity of different items. As a simple example, suppose we wish to measure prejudice against women. To do this, we create a set of 20 statements, each of which reflects that prejudice. One of the items might be "Women can't drive as well as men." Another might be "Women shouldn't be allowed to vote." Likert's scaling technique would demonstrate the difference in intensity between these items as well as pegging the intensity of the other 18 statements.

Let's suppose we ask a sample of people to agree or disagree with each of the 20 statements.

Simply giving one point for each of the indicators of prejudice against women would yield the possibility of index scores ranging from 0 to 20. A **Likert scale** goes one step beyond that and calculates the average index score for those agreeing with each of the individual statements. Let's say that all those who agreed that women are poorer drivers than men had an average index score of 1.5 (out of a possible 20). Those who agreed that women should be denied the right to vote might have an average index score of, say, 19.5—indicating the greater degree of prejudice reflected in that response.

As a result of this item analysis, respondents could be rescored to form a scale: 1.5 points for agreeing that women are poorer drivers, 19.5 points for saying women shouldn't vote, and points for other responses reflecting how those items related to the initial, simple index. If those who disagreed with the statement "I might vote for a woman for president" had an average index score of 15, then the scale would give 15 points to people disagreeing with that statement.

In practice, Likert scaling is seldom used today. I don't know why; maybe it seems too complex. The item format devised by Likert, however, is one of the most commonly used formats in contemporary questionnaire design. Typically, it is now used in the creation of simple indexes. With, say, five response categories, scores of 0 to 4 or 1 to 5 might be assigned, taking the direction of the items into account (for example, assign a score of 5 to "strongly agree" for positive items and to "strongly disagree" for negative items). Each respondent would then be assigned an overall score representing the summation of the scores he or she received for responses to the individual items.

Semantic Differential

Like the Likert format, the **semantic differential** asks respondents to a questionnaire to choose between two opposite positions by using qualifiers to bridge the distance between the two opposites. Here's how it works.

Suppose you're evaluating the effectiveness of a new music-appreciation lecture on subjects' appreciation of music. As a part of your study, you want to play some musical selections and have the subjects report their feelings about them. A good way to tap those feelings would be to use a semantic differential format.

To begin, you must determine the dimensions along which subjects should judge each selection. Then you need to find two opposite terms, representing the polar extremes along each dimension. Let's suppose one dimension that interests you is simply whether subjects enjoyed the piece or not. Two opposite terms in this case could be "enjoyable" and "unenjoyable." Similarly, you might want to know whether they regarded the individual selections as "complex" or "simple," "harmonic" or "discordant," and so forth.

Once you have determined the relevant dimensions and have found terms to represent the extremes of each, you might prepare a rating sheet each subject would complete for each piece of music. Figure 6-5 shows what it might look like.

On each line of the rating sheet, the subject would indicate how he or she felt about the piece of music: whether it was enjoyable or unenjoyable, for example, and whether it was "somewhat" that way or "very much" so. To avoid creating a biased

Likert scale A type of composite measure developed by Rensis Likert in an attempt to improve the levels of measurement in social research through the use of standardized response categories in survey questionnaires to determine the relative intensity of different items. Likert items are those using such response categories as strongly agree, agree, disagree, and strongly disagree. Such items may be used in the construction of true Likert scales as well as other types of composite measures.

semantic differential A questionnaire format in which the respondent is asked to rate something in terms of two, opposite adjectives (e.g., rate textbooks as "boring" or "exciting"), using qualifiers such as "very," "somewhat," "neither," "somewhat," and "very" to bridge the distance between the two opposites.

	Very Much	Somewhat	Neither	Somewhat	Very Much	
Enjoyable	☐	☐	☐	☐	☐	Unenjoyable
Simple	☐	☐	☐	☐	☐	Complex
Discordant	☐	☐	☐	☐	☐	Harmonic
Traditional	☐	☐	☐	☐	☐	Modern

FIGURE 6-5

Semantic Differential: Feelings about Musical Selections. The semantic differential asks respondents to describe something or someone in terms of opposing adjectives.

pattern of responses to such items, it's a good idea to vary the placement of terms that are likely to be related to each other. Notice, for example, that "discordant" and "traditional" are on the left side of the sheet, with "harmonic" and "modern" on the right. Most likely, those selections scored as "discordant" would also be scored as "modern" as opposed to "traditional."

Both the Likert and semantic differential formats have a greater rigor and structure than other question formats do. As I indicated earlier, these formats produce data suitable to both indexing and scaling.

Guttman Scaling

Researchers today often use the scale developed by Louis Guttman. Like Bogardus, Thurstone, and Likert scaling, Guttman scaling is based on the fact that some items under consideration may prove to be more-extreme indicators of the variable than others. Here's an example to illustrate this pattern.

In the earlier example of measuring scientific orientation among medical school faculty members, you'll recall that a simple index was constructed. As it happens, however, the three items included in the index essentially form a Guttman scale.

The construction of a **Guttman scale** begins with some of the same steps that initiate index construction. You begin by examining the face validity of items available for analysis. Then, you examine the bivariate and perhaps multivariate relations among those items. In scale construction, however, you also look for relatively "hard" and "easy" indicators of the variable being examined.

Guttman scale A type of composite measure used to summarize several discrete observations and to represent some more-general variable.

Earlier, when we talked about attitudes regarding a woman's right to have an abortion, we discussed several conditions that can affect people's opinions: whether the woman is married, whether her health is endangered, and so forth. These differing conditions provide an excellent illustration of Guttman scaling.

Here are the percentages of the people in the 2000 GSS sample who supported a woman's right to an abortion, under three different conditions:

Woman's health is seriously endangered	89%
Pregnant as a result of rape	81%
Woman is not married	39%

The different percentages supporting abortion under the three conditions suggest something about the different levels of support that each item indicates. For example, if someone supports abortion when the mother's life is seriously endangered, that's not a very strong indicator of general support for abortion, because almost everyone agreed with that. Supporting abortion for unmarried women seems a much stronger indicator of support for abortion in general—fewer than half the sample took that position.

Guttman scaling is based on the idea that anyone who gives a strong indicator of some variable will also give the weaker indicators. In this case, we

TABLE 6-2

Scaling Support for Choice of Abortion

	Women's Health	*Result of Rape*	*Woman Unmarried*	*Number of Cases*
	+	+	+	677
	+	+	−	607
Scale Types	+	−	−	165
	−	−	−	147
				Total = 1,596
	−	+	−	42
	+	−	+	5
Mixed Types	−	−	+	2
	−	+	+	4
				Total = 53

+ = favors woman's right to choose; − = opposes woman's right to choose

would assume that anyone who supported abortion for unmarried women would also support it in the case of rape or of the woman's health being threatened. Table 6-2 tests this assumption by presenting the number of respondents who gave each of the possible response patterns.

The first four response patterns in the table compose what we would call the *scale types*: those patterns that form a scalar structure. Following those respondents who supported abortion under all three conditions (line 1), we see (line 2) that those with only two pro-choice responses have chosen the two easier ones; those with only one such response (line 3) chose the easiest of the three (the woman's health being endangered). And finally, there are some respondents who opposed abortion in all three circumstances (line 4).

The second part of the table presents those response patterns that violate the scalar structure of the items. The most radical departures from the scalar structure are the last two response patterns: those who accepted only the hardest item and those who rejected only the easiest one.

The final column in the table indicates the number of survey respondents who gave each of the response patterns. The great majority (1,596, or 99 percent) fit into one of the scale types. The presence of mixed types, however, indicates that the items do not form a perfect Guttman scale. (It would be extremely rare for such data to form a Guttman scale *perfectly*.)

Recall at this point that one of the chief functions of scaling is efficient data reduction. Scales provide a technique for presenting data in a summary form while maintaining as much of the original information as possible. When the scientific orientation items were formed into an index in our earlier discussion, respondents were given one point for each scientific response they gave. If these same three items were scored as a Guttman scale, some respondents would be assigned scale scores that would permit the most accurate reproduction of their original responses to all three items.

In the present example of attitudes regarding abortion, respondents fitting into the scale types would receive the same scores as would be assigned in the construction of an index. Persons selecting all three pro-choice responses (+ + +) would still be scored 3, those who selected pro-choice responses to the two easier items and were opposed on the hardest item (+ + −) would be scored 2, and so on. For each of the four scale types we could predict accurately all the actual responses given by all the respondents based on their scores.

The mixed types in the table present a problem, however. The first mixed type (− + −) was scored 1 on the index to indicate only one pro-choice response. But, if 1 were assigned as a scale score, we would predict that the 42 respondents in this group had chosen only the easiest item (approving abortion when the woman's life was endangered), and we would be making two errors for each such respondent: thinking their response pattern was (+ − −) instead of (− + −). Scale scores are assigned, therefore, with the aim of minimizing the errors that would be made in reconstructing the original responses.

Table 6-3 illustrates the index and scale scores that would be assigned to each of the response patterns in our example. Note that one error is made for each respondent in the mixed types. This is the

TABLE 6-3

Index and Scale Scores

	Response Pattern	Number of Cases	Index Scores	Scale Scores	Total Scale Errors
Scale Types	+ + +	677	3	3	0
	+ + −	607	2	2	0
	+ − −	165	1	1	0
	− − −	147	0	0	0
Mixed Types	− + −	42	1	2	42
	+ − +	5	2	3	5
	− − +	2	1	0	2
	− + +	4	2	3	4

Total Scale errors = 53

$$\text{Coefficient of reproducibility} = 1 - \frac{\text{number of errors}}{\text{number of guesses}}$$

$$= 1 - \frac{53}{1{,}649 \times 3} = 1 - \frac{53}{4{,}947}$$

$$= 0.989 = 98.9\%$$

This table presents one common method for scoring mixed types, but you should be advised that other methods are also used.

minimum we can hope for in a mixed-type pattern. In the first mixed type, for example, we would erroneously predict a pro-choice response to the easiest item for each of the 42 respondents in this group, making a total of 42 errors.

The extent to which a set of empirical responses form a Guttman scale is determined by the accuracy with which the original responses can be reconstructed from the scale scores. For each of the 1,649 respondents in this example, we'll predict three questionnaire responses, for a total of 4,947 predictions. Table 6-3 indicates that we'll make 53 errors using the scale scores assigned. The percentage of correct predictions is called the *coefficient of reproducibility:* the percentage of original responses that could be reproduced by knowing the scale scores used to summarize them. In the present example, the coefficient of reproducibility is 4,894/4,947, or 98.9 percent.

Except for the case of perfect (100 percent) reproducibility, there is no way of saying that a set of items does or does not form a Guttman scale in any absolute sense. Virtually all sets of such items approximate a scale. As a general guideline, however, coefficients of 90 or 95 percent are the commonly used standards. If the observed reproducibility exceeds the level you've set, you'll probably decide to score and use the items as a scale.

The decision concerning criteria in this regard is, of course, arbitrary. Moreover, a high degree of reproducibility does not insure that the scale constructed in fact measures the concept under consideration. What it does is increase confidence that all the component items measure *the same thing*. Also, you should realize that a high coefficient of reproducibility is most likely when few items are involved.

One concluding remark with regard to Guttman scaling: It's based on the structure observed among the actual data under examination. This is an important point that is often misunderstood. It does not make sense to say that a set of questionnaire items (perhaps developed and used by a previous researcher) constitutes a Guttman scale. Rather, we can say only that they form a scale within a given body of data being analyzed. Scalability, then, is a sample-dependent, empirical matter. Although a set of items may form a Guttman scale among one sample of survey respondents, for example, there is no guarantee that this set will form such a scale among another sample. In this sense, then, a set of questionnaire items in and of itself never forms a scale, but a set of empirical observations may.

This concludes our discussion of indexing and scaling. Like indexes, scales are composite measures of a variable, typically broadening the meaning of the variable beyond what might be captured by a single indicator. Both scales and indexes seek to measure variables at the ordinal level of measurement. Unlike indexes, however, scales take advantage of any intensity structure that may be present among the individual indicators. To the extent that such an intensity structure is found and the data from the people or other units of analysis comply with the logic of that intensity structure, we can have confidence that we have created an ordinal measure.

Typologies

We conclude this chapter with a short discussion of typology construction and analysis. Recall that indexes and scales are constructed to provide ordinal measures of given variables. We attempt to assign index or scale scores to cases in such a way as to indicate a rising degree of prejudice, religiosity, conservatism, and so forth. In such cases, we're dealing with single dimensions.

Often, however, the researcher wishes to summarize the intersection of two or more variables, thereby creating a set of categories or types—a nominal variable—called a **typology**. You may, for example, wish to examine the political orientations of newspapers separately in terms of domestic issues and foreign policy. The fourfold presentation in Table 6-4 describes such a typology.

Newspapers in cell A of the table are conservative on both foreign policy and domestic policy; those in cell D are liberal on both. Those in cells B and C are conservative on one and liberal on the other.

Frequently, you arrive at a typology in the course of an attempt to construct an index or scale. The items that you felt represented a single variable appear to represent two. We might have been attempting to construct a single index of political orientations for newspapers but discovered—empirically—that foreign and domestic politics had to be kept separate.

In any event, you should be warned against a difficulty inherent in typological analysis. Whenever the typology is used as the independent variable, there will probably be no problem. In the preceding example, you might compute the percentages of newspapers in each cell that normally endorse Democratic candidates; you could then easily examine the effects of both foreign and domestic policies on political endorsements.

It's extremely difficult, however, to analyze a typology as a dependent variable. If you want to discover *why* newspapers fall into the different cells of typology, you're in trouble. That becomes apparent when we consider the ways you might construct and read your tables. Assume, for example, that you want to examine the effects of community size on political policies. With a single dimension, you could easily determine the percentages of rural and urban newspapers that were scored conservative and liberal on your index or scale.

TABLE 6-4
A Political Typology of Newspapers

		Foreign Policy	
		Conservative	*Liberal*
Domestic Policy	Conservative	A	B
	Liberal	C	D

With a typology, however, you would have to present the distribution of the urban newspapers in your sample among types A, B, C, and D. Then you would repeat the procedure for the rural ones in the sample and compare the two distributions. Let's suppose that 80 percent of the rural newspapers are scored as type A (conservative on both dimensions), compared with 30 percent of the urban ones. Moreover, suppose that only 5 percent of the rural newspapers are scored as type B (conservative only on domestic issues), compared with 40 percent of the urban ones. It would be incorrect to conclude from an examination of type B that urban newspapers are more conservative on domestic issues than rural ones are, because 85 percent of the rural newspapers, compared with 70 percent of the urban ones, have this characteristic. The relative sparsity of rural newspapers in type B is due to their concentration in type A. It should be apparent that an interpretation of such data would be very difficult for anything other than description.

In reality, you'd probably examine two such dimensions separately, especially if the dependent variable has more categories of responses than the given example does.

Don't think that typologies should always be avoided in social research; often they provide the

typology The classification (typically nominal) of observations in terms of their attributes on two or more variables. The classification of newspapers as liberal-urban, liberal-rural, conservative-urban, or conservative-rural would be an example.

most appropriate device for understanding the data. To examine the pro-life orientation in depth, for example, you might create a typology involving both abortion and capital punishment. Libertarianism could be seen in terms of both economic and social permissiveness. You've now been warned, however, against the special difficulties involved in using typologies as dependent variables.

MAIN POINTS

Introduction

- Single indicators of variables seldom capture all the dimensions of a concept, have sufficiently clear validity to warrant their use, or permit the desired range of variation to allow ordinal rankings. Composite measures, such as scales and indexes, solve these problems by including several indicators of a variable in one summary measure.

Indexes versus Scales

- Although both indexes and scales are intended as ordinal measures of variables, scales typically satisfy this intention better than indexes do.
- Whereas indexes are based on the simple cumulation of indicators of a variable, scales take advantage of any logical or empirical intensity structures that exist among a variable's indicators.

Index Construction

- The principal steps in constructing an index include selecting possible items, examining their empirical relationships, scoring the index, and validating it.
- Criteria of item selection include face validity, unidimensionality, the degree of specificity with which a dimension is to be measured, and the amount of variance provided by the items.
- If different items are indeed indicators of the same variable, then they should be related empirically to one another. In constructing an index, the researcher needs to examine bivariate and multivariate relationships among the items.
- Index scoring involves deciding the desirable range of scores and determining whether items will have equal or different weights.
- There are various techniques that allow items to be used in an index in spite of missing data.
- Item analysis is a type of internal validation, based on the relationship between individual items in the composite measure and the measure itself. External validation refers to the relationships between the composite measure and other indicators of the variable—indicators not included in the measure.

Scale Construction

- Four types of scaling techniques are represented by the Bogardus social distance scale, a device for measuring the varying degrees to which a person would be willing to associate with a given class of people; Thurstone scaling, a technique that uses judges to determine the intensities of different indicators; Likert scaling, a measurement technique based on the use of standardized response categories; and Guttman scaling, a method of discovering and using the empirical intensity structure among several indicators of a given variable. Guttman scaling is probably the most popular scaling technique in social research today.
- The semantic differential is a question format that asks respondents to make ratings that lie between two extremes, such as "very positive" and "very negative."

Typologies

- A typology is a nominal composite measure often used in social research. Typologies may be used effectively as independent variables, but interpretation is difficult when they are used as dependent variables.

KEY TERMS

The following terms are defined in context in the chapter and at the bottom of the page where the term is introduced, as well as in the comprehensive glossary at the back of the book.

Bogardus social distance scale
external validation
Guttman scale
index
item analysis
Likert scale
scale
semantic differential
Thurstone scale
typology

REVIEW QUESTIONS AND EXERCISES

1. In your own words, describe the difference between an index and a scale.
2. Suppose you wanted to create an index for rating the quality of colleges and universities. Name three data items that might be included in such an index.
3. Make up three questionnaire items that measure attitudes toward nuclear power and that would probably form a Guttman scale.
4. Construct a typology of pro-life attitudes as discussed in the chapter.
5. Economists often use indexes to measure economic variables, such as the cost of living. Go to the Bureau of Labor Statistics (http://www.bls.gov) and find the Consumer Price Index survey. What are some of the dimensions of living costs included in this measure?

ADDITIONAL READINGS

Anderson, Andy B., Alexander Basilevsky, and Derek P. J. Hum. 1983. "Measurement: Theory and Techniques." Pp. 231–87 in *Handbook of Survey Research*, edited by Peter H. Rossi, James D. Wright, and Andy B. Anderson. New York: Academic Press. The logic of measurement is analyzed in the context of composite measures.

Bobo, Lawrence, and Frederick C. Licari. 1989. "Education and Political Tolerance: Testing the Effects of Cognitive Sophistication and Target Group Effect." *Public Opinion Quarterly* 53:285–308. The authors use a variety of techniques for determining how best to measure tolerance toward different groups in society.

Indrayan, A., M. J. Wysocki, A. Chawla, R. Kumar, and N. Singh. 1999. "Three-Decade Trend in Human Development Index in India and Its Major States." *Social Indicators Research* 46 (1): 91–120. The authors use several human development indexes to compare the status of different states in India.

Lazarsfeld, Paul, Ann Pasanella, and Morris Rosenberg, eds. 1972. *Continuities in the Language of Social Research*. New York: Free Press. See especially Section 1. An excellent collection of conceptual discussions and concrete illustrations. The construction of composite measures is presented within the more general area of conceptualization and measurement.

McIver, John P., and Edward G. Carmines. 1981. *Unidimensional Scaling*. Newbury Park, CA: Sage. Here's an excellent way to pursue Thurstone, Likert, and Guttman scaling in further depth.

Miller, Delbert. 1991. *Handbook of Research Design and Social Measurement*. Newbury Park, CA: Sage. An excellent compilation of frequently used and semistandardized scales. The many illustrations reported in Part 4 of the Miller book may be directly adaptable to studies or at least suggestive of modified measures. Studying the several illustrations, moreover, may also give you a better understanding of the logic of composite measures in general.

SPSS EXERCISES

See the booklet that accompanies your text for exercises using SPSS (Statistical Package for the Social Sciences). There are exercises offered for each chapter, and you'll also find a detailed primer on using SPSS.

Online Study Resources

SociologyNow™: Research Methods

1. Before you do your final review of the chapter, take the *SociologyNow: Research Methods* diagnostic quiz to help identify the areas on which you should concentrate. You'll find information on this online tool, as well as instructions on how to access all of its great resources, in the front of the book.
2. As you review, take advantage of the *Sociology Now: Research Methods* customized study plan, based on your quiz results. Use this study plan with its interactive exercises and other resources to master the material.

3. When you're finished with your review, take the posttest to confirm that you're ready to move on to the next chapter.

WEBSITE FOR *THE PRACTICE OF SOCIAL RESEARCH* 11TH EDITION

Go to your book's website at http://sociology.wadsworth.com/babbie_practice11e for tools to aid you in studying for your exams. You'll find *Tutorial Quizzes* with feedback, *Internet Exercises*, *Flashcards*, and *Chapter Tutorials*, as well as *Extended Projects*, *InfoTrac College Edition* search terms, *Social Research in Cyberspace*, *GSS Data*, *Web Links*, and primers for using various data-analysis software such as SPSS and NVivo.

WEB LINKS FOR THIS CHAPTER

Please realize that the Internet is an evolving entity, subject to change. Nevertheless, these few websites should be fairly stable. Also, check your book's website for even more *Web Links*. These websites, current at the time of this book's publication, provide opportunities to learn about indexes, scales, and typologies.

Bureau of Labor Statistics, Measurement Issues in the Consumer Price Index
http://www.bls.gov/cpi/cpigm697.htm
The federal government's Consumer Price Index (CPI) is one of those composite measures that affects many people's lives—determining cost-of-living increases, for example. This site discusses some aspects of the measure.

Arizona State University, Reliability and Validity
http://seamonkey.ed.asu.edu/~alex/teaching/assessment/reliability.html
Here you'll find an extensive discussion of these two aspects of measurement quality.

Thomas O'Connor, "Scales and Indexes"
http://faculty.ncwc.edu/toconnor/308/308lect05.htm
This web page has an excellent discussion of scales and indexes in general, provides illustrative examples, and also gives hot links useful for pursuing the topic.