Modes of Observation: Quantitative and Qualitative

Chapter 11 discusses three forms of unobtrusive data collection that take advantage of some of the data available all around us. For example, content analysis is a method of collecting social data through carefully specifying and counting social artifacts such as books, songs, speeches, and paintings. Without making any personal contact with people, you can use this method to examine a wide variety of social phenomena. The analysis of existing statistics offers another way of studying people without having to talk to them. Governments and a variety of private organizations regularly compile great masses of data, which you often can use with little or no modification to answer properly posed questions. Finally, historical documents are a valuable resource for social science analysis.

Chapter 12, on evaluation research, looks at a rapidly growing subfield in social science involving the application of experimental and quasi-experimental models to the testing of social interventions in real life. You might use evaluation research, for example, to test the effectiveness of a drug rehabilitation program or the efficiency of a new school cafeteria. In the same chapter, we'll look briefly at social indicators as a way of assessing broader social processes.
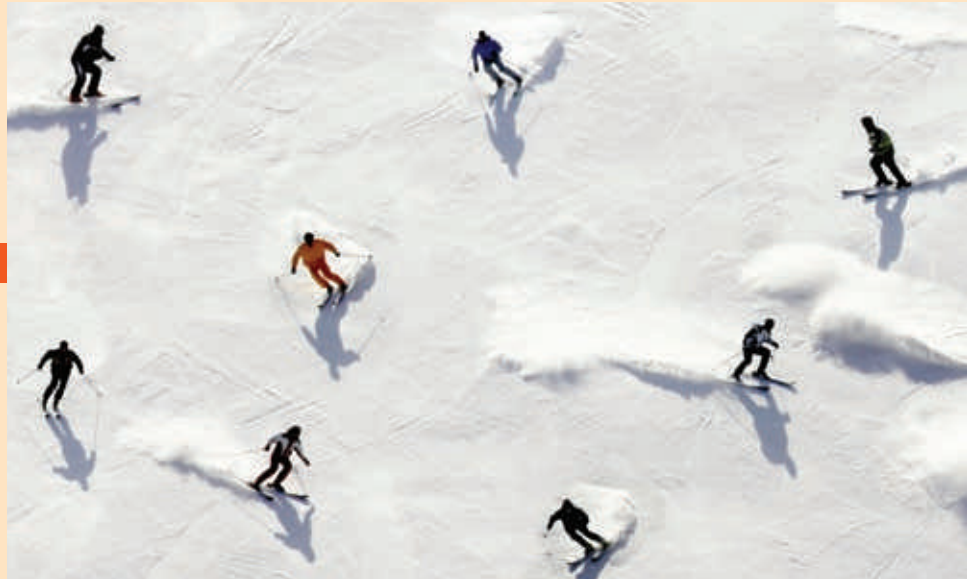
Before we turn to the actual descriptions of these research methods, two points should be made. First, you'll probably discover that you've been using these scientific methods casually in your daily life for as long as you can remember. You use some form of field re-search every day. You employ a crude form of content analysis every time you judge an author's motivation from her or his writings. You perform casual experiments every day. Part 3 will show you how to improve your use of these methods so as to avoid certain pitfalls.

Second, none of the data-collection methods described in these chapters is appropriate to all research topics and situations. I give you some ideas, early in each chapter, regarding when a given method might be appro-priate. Still, I could never anticipate all the research topics that may one day interest you. As a general guideline, you should always use a variety of techniques in the study of any topic. Because each method has its weaknesses, the use of several methods can help fill any gaps; if the dif-ferent, independent approaches to the topic all yield the same conclusion, you've achieved a form of replication.

Experiments

CHAPTER OVERVIEW

An experiment is a mode of observation that enables researchers to probe causal relationships. Many experiments in social research are conducted under the controlled conditions of a laboratory, but experimenters can also take advantage of natural occurrences to study the effects of events in the social world.

Introduction

This chapter addresses the controlled experiment: a research method associated more with the natural than the social sciences. We begin Part 3 with this method because the logic and basic techniques of the controlled experiment provide a useful backdrop for understanding other techniques more commonly used in social science, especially for explanatory purposes. We'll also see in this chapter some of the inventive ways social scientists have conducted experiments.

At base, experiments involve (1) taking action and (2) observing the consequences of that action. Social researchers typically select a group of subjects, do something to them, and observe the effect of what was done.

It's worth noting at the outset that we often use experiments in nonscientific inquiry. In preparing a stew, for example, we add salt, taste, add more salt, and taste again. In defusing a bomb, we clip the red wire, observe whether the bomb explodes, clip another, and . . .

We also experiment copiously in our attempts to develop generalized understandings about the world we live in. All skills are learned through experimentation: eating, walking, talking, riding a bicycle, swimming, and so forth. Through experimentation, students discover how much studying is required for academic success. Through experimentation, professors learn how much preparation is required for successful lectures. This chapter discusses how social researchers use experiments to develop generalized understandings. We'll see that, like other methods available to the social researcher, experimenting has its special strengths and weaknesses.

Topics Appropriate for Experiments

Experiments are more appropriate for some topics and research purposes than others. Experiments are especially well suited to research projects involving relatively limited and well-defined concepts and propositions. In terms of the traditional image of science, discussed earlier in this book, the experimental model is especially appropriate for hypothesis testing. Because experiments focus on determining causation, they're also better suited to explanatory than to descriptive purposes.

Let's assume, for example, that we want to discover ways of reducing prejudice against Muslims. We hypothesize that learning about the contribution of Muslims to U.S. history will reduce prejudice, and we decide to test this hypothesis experimentally. To begin, we might test a group of experimental subjects to determine their levels of prejudice against Muslims. Next, we might show them a documentary film depicting the many important ways Muslims have contributed to the scientific, literary, political, and social development of the nation. Finally, we would measure our subjects' levels of prejudice against Muslims to determine whether the film has actually reduced prejudice.

Experimentation has also been successful in the study of small-group interaction. Thus, we might bring together a small group of experimental subjects and assign them a task, such as making recommendations for popularizing car pools. We observe, then, how the group organizes itself and deals with the problem. Over the course of several such experiments, we might systematically vary the nature of the task or the rewards for handling the task successfully. By observing differences in the way groups organize themselves and operate under these varying conditions, we can learn a great deal about the nature of small-group interaction and the factors that influence it. For example, attorneys sometimes present evidence in different ways to different mock juries, to see which method is the most effective.

Political campaigns use experimental methods to determine the most effective types of communication. Different fund-raising messages are evaluated in terms of the funds actually raised.

Laboratory experiments have been used less frequently in the social sciences than in psychology and the natural sciences. Researchers Christine Horne and Michael Lovaglia (2008) argue that this has been a shortcoming in the field of criminology. They have gathered a number of

examples to reveal how laboratory experiments have contributed to understanding with regard to such topics as self-control, social influence, and the law. Horne and Lovaglia do not argue for the replacement of other methods but advocate that studies be augmented with research in laboratory settings.

Similarly, Howard Schuman (2008) details ways in which laboratory experiments can evaluate the effects of differences in question wording and question order in survey research. As we'll see in the next chapter, experienced survey researchers have found differences in public support (or nonsupport) depending on whether government programs are called "welfare" or "assistance to the poor." However, carefully designed experiments can uncover wording impacts that might not be as evident or intuitive to designers of research.

We typically think of experiments as being conducted in laboratories. Indeed, most of the examples in this chapter involve such a setting. This need not be the case, however. Increasingly, social researchers are using the Internet as a vehicle for conducting experiments. Further, sometimes we can construct what are called natural experiments: "experiments" that occur in the regular course of social events. The latter portion of this chapter deals with such research.

The Classical Experiment

In both the natural and the social sciences, the most conventional type of experiment involves three major pairs of components: (1) independent and dependent variables, (2) pretesting and posttesting, and (3) experimental and control groups. This section looks at each of these components and the way they're put together in the execution of the experiment.

Independent and Dependent Variables

Essentially, an experiment examines the effect of an independent variable on a dependent variable. Typically, the independent variable takes the form of an experimental stimulus, which is either present or absent. That is, the stimulus is a dichotomous variable, having two attributes, present or not present. In this typical model, the experimenter compares what happens when the stimulus is present to what happens when it is not.

In the example concerning prejudice against Muslims, *prejudice* is the dependent variable and *exposure to Muslim history* is the independent variable. The researcher's hypothesis suggests that prejudice depends, in part, on a lack of knowledge of Muslim history. The purpose of the experiment is to test the validity of this hypothesis by presenting some subjects with an appropriate stimulus, such as a documentary film. In other terms, the independent variable is the cause and the dependent variable is the effect. Thus, we might say that watching the film caused a change in prejudice or that reduced prejudice was an effect of watching the film.

The independent and dependent variables appropriate for experimentation are nearly limitless. Moreover, a given variable might serve as an independent variable in one experiment and as a dependent variable in another. For example, *prejudice* is the dependent variable in our example, but it might be the independent variable in an experiment examining the effect of prejudice on voting behavior.

To be used in an experiment, both independent and dependent variables must be operationally defined. Such operational definitions might involve a variety of observation methods. Responses to a questionnaire, for example, might be the basis for defining prejudice. Speaking to or ignoring Muslims, or agreeing or disagreeing with them, might be elements in the operational definition of interaction with Muslims in a small-group setting.

Conventionally, in the experimental model, dependent and independent variables must be operationally defined before the experiment begins. However, as you'll see in connection with survey research and other methods, it's sometimes appropriate to make a wide variety of observations during data collection and then determine the most useful operational definitions of variables during later analyses. Ultimately, however, experimentation, like other quantitative methods, requires specific and standardized measurements and observations.

Pretesting and Posttesting

In the simplest experimental design, subjects are measured in terms of a dependent variable

(**pretesting**), exposed to a stimulus representing an independent variable, and then remeasured in terms of the dependent variable (**posttesting**). Any differences between the first and last measurements on the dependent variable are then attributed to the independent variable.

In the example of prejudice and exposure to Muslim history, we'd begin by pretesting the extent of prejudice among our experimental subjects. Using a questionnaire asking about attitudes toward Muslims, for example, we could measure both the extent of prejudice exhibited by each individual subject and the average prejudice level of the whole group. After exposing the subjects to the Muslim history film, we could administer the same questionnaire again. Responses given in this posttest would permit us to measure the later extent of prejudice for each subject and the average prejudice level of the group as a whole. If we discovered a lower level of prejudice during the second administration of the questionnaire, we might conclude that the film had indeed reduced prejudice.

In the experimental examination of attitudes such as prejudice, we face a special practical problem relating to validity. As you may already have imagined, the subjects might respond differently to the questionnaires the second time even if their attitudes remain unchanged. During the first administration of the questionnaire, the subjects might be unaware of its purpose. By the second measurement, they might have figured out that the researchers were interested in measuring their prejudice. Because no one wishes to seem prejudiced, the subjects might "clean up" their answers the second time around. Thus, the film would seem to have reduced prejudice although, in fact, it had not.

This is an example of a more general problem that plagues many forms of social research: The very act of studying something may change it. The techniques for dealing with this problem in the context of experimentation will be discussed in various places throughout the chapter. The first technique involves the use of control groups.

Experimental and Control Groups

Laboratory experiments seldom, if ever, involve only the observation of an **experimental group** to which a stimulus has been administered. In
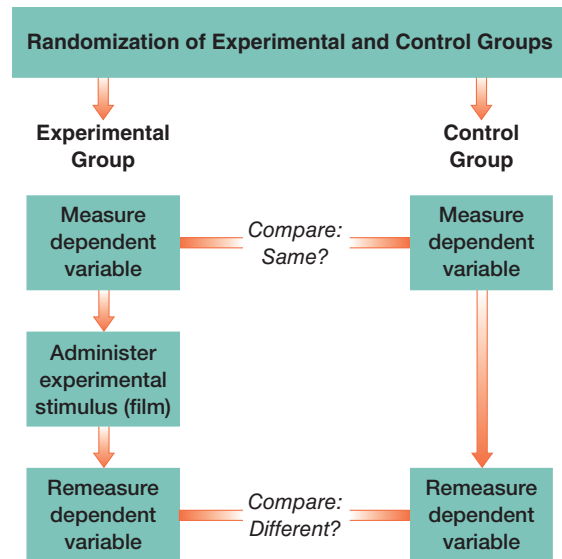

FIGURE 8-1

Diagram of Basic Experimental Design. The fundamental purpose of an experiment is to isolate the possible effect of an independent variable (called the *stimulus* in experiments) on a dependent variable. Members of the experimental group(s) are exposed to the stimulus, whereas those in the control group(s) are not.

© Cengage Learning®

addition, the researchers also observe a **control group**, which does not receive the experimental stimulus.

In the example of prejudice and Muslim history, we might examine two groups of subjects. To begin, we give each group a questionnaire designed to measure their prejudice against Muslims. Then we show the film to only the experimental group. Finally, we administer a posttest of prejudice to both groups. Figure 8-1 illustrates this basic experimental design.

pretesting The measurement of a dependent variable among subjects.

posttesting The remeasurement of a dependent variable among subjects after they've been exposed to an independent variable.

experimental group In experimentation, a group of subjects to whom an experimental stimulus is administered.

control group In experimentation, a group of subjects to whom no experimental stimulus is administered and who should resemble the experimental group in all other respects. The comparison of the control group and the experimental group at the end of the experiment points to the effect of the experimental stimulus.

Using a control group allows the researcher to detect any effects of the experiment itself. If the posttest shows that the overall level of prejudice exhibited by the control group has dropped as much as that of the experimental group, then the apparent reduction in prejudice must be a function of the experiment or of some external factor rather than a function of the film. If, on the other hand, prejudice is reduced only in the experimental group, this reduction would seem to be a consequence of exposure to the film, because that's the only difference between the two groups. Alternatively, if prejudice is reduced in both groups but to a greater degree in the experimental group than in the control group, that, too, would be grounds for assuming that the film reduced prejudice.

The need for control groups in social research became clear in connection with a series of studies of employee satisfaction conducted by F. J. Roethlisberger and W. J. Dickson (1939) in the late 1920s and early 1930s. These two researchers were interested in discovering what changes in working conditions would improve employee satisfaction and productivity. To pursue this objective, they studied working conditions in the telephone "bank wiring room" of the Western Electric Works in the Chicago suburb of Hawthorne, Illinois.

To the researchers' great satisfaction, they discovered that improving the working conditions increased satisfaction and productivity consistently. As the workroom was brightened up through better lighting, for example, productivity went up. When lighting was further improved, productivity went up again.

To further substantiate their scientific conclusion, the researchers then *dimmed* the lights. Whoops—productivity improved again!

At this point it became evident that the wiring-room workers were responding more to the attention given them by the researchers than to improved working conditions. As a result of this phenomenon, often called the *Hawthorne effect,* social researchers have become more sensitive to and cautious about the possible effects of experiments themselves. In the wiring-room study, the use of a proper control group—one that was studied intensively without any other changes in the working conditions—would have pointed to the presence of this effect.

The need for control groups in experimentation has been nowhere more evident than in medical research. Time and again, patients who participate in medical experiments have appeared to improve, but it has been unclear how much of the improvement has come from the experimental treatment and how much from the experiment. In testing the effects of new drugs, then, medical researchers frequently administer a *placebo*—a "drug" with no relevant effect, such as sugar pills—to a control group. Thus, the control-group patients believe that they, like the experimental group, are receiving an experimental drug. Often, they improve. If the new drug is effective, however, those receiving the actual drug will improve more than those receiving the placebo.

In social science experiments, control groups guard against not only the effects of the experiments themselves but also the effects of any events outside the laboratory during the experiments. In the example of the study of prejudice, suppose that a popular Muslim leader is assassinated in the middle of, say, a weeklong experiment. Such an event may very well horrify the experimental subjects, requiring them to examine their own attitudes toward Muslims, with the result of reduced prejudice. Because such an effect should happen about equally for members of the control and experimental groups, a greater reduction of prejudice among the experimental group would, again, point to the impact of the experimental stimulus: the documentary film.

Sometimes an experimental design requires more than one experimental or control group. In the case of the documentary film, for example, we might also want to examine the impact of reading a book about Muslim history. In that case, we might have one group see the film and read the book, another group only see the movie, still another group only read the book, and the control group do neither. With this kind of design, we could determine the impact of each stimulus separately, as well as their combined effect.

The Double-Blind Experiment

Like patients who improve when they merely think they're receiving a new drug, sometimes experimenters tend to prejudge results. In

medical research, the experimenters may be more likely to "observe" improvements among patients receiving the experimental drug than among those receiving the placebo. (This would be most likely, perhaps, for the researcher who developed the drug.) A **double-blind experiment** eliminates this possibility, because in this design neither the subjects nor the experimenters know which is the experimental group and which is the control. In the medical case, those researchers who were responsible for administering the drug and for noting improvements would not be told which subjects were receiving the drug and which the placebo. Conversely, the researcher who knew which subjects were in which group would not administer the experiment.

In social science experiments, as in medical experiments, the danger of experimenter bias is further reduced to the extent that the operational definitions of the dependent variables are clear and precise. Thus, medical researchers would be less likely to unconsciously bias their reading of a patient's temperature than they would be to bias their assessment of how lethargic the patient was. For the same reason, the small-group researcher would be less likely to misperceive which subject spoke, or to whom he or she spoke, than whether the subject's comments sounded cooperative or competitive, a more subjective judgment that's difficult to define in precise behavioral terms.

The role of the placebo may be more complex than you think, according to a 2010 medical experiment on irritable bowel syndrome. One group of sufferers was given pills in a bottle marked "Placebo" and it was explained that a placebo, sometimes called a sugar pill, contained no active ingredients. Subjects were told that people sometimes seemed to benefit from the placebos. A control group was given no treatment at all. After 21 days the placebo group had improved significantly, while the control group had not.

This study is further complicated, however, by the fact that those receiving the placebo pills also received examinations and counseling sessions, while the control group received no attention at all. Perhaps, as the researchers acknowledge, the positive results were produced by the comprehensive treatment package, not by

the placebo pills alone. Also, they note, the measures of improvement were self-assessments. It is possible that physiological measurements might have shown no improvement. But, to complicate matters further, isn't "feeling better" the goal of such treatments?

Selecting Subjects

In Chapter 7 we discussed the logic of sampling, which involves selecting a sample that is representative of some population. Similar considerations apply to experiments. Because most social researchers work in colleges and universities, it seems likely that research laboratory experiments would be conducted with college undergraduates as subjects. Typically, the experimenter asks students enrolled in his or her classes to participate in experiments or advertises for subjects in a college newspaper. Subjects may or may not be paid for participating in such experiments (recall also from Chapter 3 the ethical issues involved in asking students to participate in such studies).

In relation to the norm of generalizability in science, this tendency clearly represents a potential defect in social research. Simply put, college undergraduates are not typical of the public at large. There is a danger, therefore, that we may learn much about the attitudes and actions of college undergraduates but not about social attitudes and actions in general.

However, this potential defect is less significant in explanatory research than in descriptive research. True, having noted the level of prejudice among a group of college undergraduates in our pretesting, we would have little confidence that the same level existed among the public at large. On the other hand, if we found that a documentary film reduced whatever level of prejudice existed among those undergraduates, we would have more confidence—without being certain—that it would have a comparable effect in the community at large. Social processes

double-blind experiment An experimental design in which neither the subjects nor the experimenters know which is the experimental group and which is the control.

and patterns of causal relationships appear to be more generalizable and more stable than specific characteristics such as an individual's level of prejudice.

This problem of generalizing from students isn't always seen as problematic, as Jerome Taylor reports in a commentary on the research into the common cold, a disease he traces back to ancient Egypt. This elusive illness only attacks humans and chimpanzees, so you can probably guess how medical researchers have selected subjects. However, you might be wrong.

> Chimpanzees were too expensive to import en masse, so during the first half of the 20th century British scientists began looking into how the common cold worked by conducting experiments on medical students at St Bartholomew's Hospital in London.
> *(Taylor 2008)*

Aside from the question of generalizability, the cardinal rule of subject selection in experimentation concerns the comparability of experimental and control groups. Ideally, the control group represents what the experimental group would be like if it had *not* been exposed to the experimental stimulus. The logic of experiments requires, therefore, that experimental and control groups be as similar as possible. There are several ways to accomplish this.

Probability Sampling

The discussions of the logic and techniques of probability sampling in Chapter 7 provide one method for selecting two groups of people that are similar to each other. Beginning with a sampling frame composed of all the people in the population under study, the researcher might select two probability samples. If these samples each resemble the total population from which they're selected, they'll also resemble each other.

Recall also, however, that the degree of resemblance (representativeness) achieved by probability sampling is largely a function of the sample size. As a general guideline, probability

samples of less than 100 are not likely to be terribly representative, and social science experiments seldom involve that many subjects in either experimental or control groups. As a result, then, probability sampling is seldom used in experiments to select subjects from a larger population. Researchers do, however, use the logic of random selection when they assign subjects to groups.

Randomization

Having recruited, by whatever means, a total group of subjects, the experimenter may randomly assign those subjects to either the experimental or the control group. The researcher might accomplish such **randomization** by numbering all of the subjects serially and selecting numbers by means of a random number table. Alternatively, the experimenter might assign the odd-numbered subjects to the experimental group and the even-numbered subjects to the control group.

Let's return again to the basic concept of probability sampling. For example, if we use a newspaper advertisement to recruit a total of 40 subjects, there's no reason to believe that these 40 subjects represent the entire population from which they've been drawn. Nor can we assume that the 20 subjects randomly assigned to the experimental group represent that larger population. We can have greater confidence, however, that the 20 subjects randomly assigned to the experimental group will be reasonably similar to the 20 assigned to the control group.

Following the logic of our earlier discussions of sampling, we can see our 40 subjects as a population from which we select two probability samples—each consisting of half the population. Because each sample reflects the characteristics of the total population, the two samples will mirror each other.

As we saw in Chapter 7, our assumption of similarity in the two groups depends in part on the number of subjects involved. In the extreme case, if we recruited only two subjects and assigned, by the flip of a coin, one as the experimental subject and one as the control, there would be no reason to assume that the two subjects are similar to each other. With larger numbers of subjects, however, randomization makes good sense.

randomization A technique for assigning experimental subjects to experimental and control groups randomly.
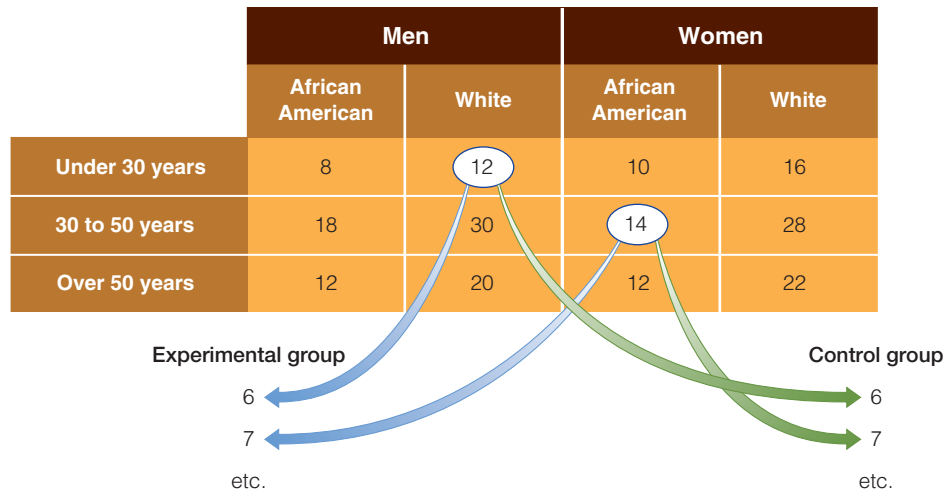
	Men		Women	
	African American	White	African American	White
Under 30 years	8	12	10	16
30 to 50 years	18	30	14	28
Over 50 years	12	20	12	22

Experimental group

6

7

etc.

Control group

6

7

etc.

FIGURE 8-2

Quota Matrix Illustration. Sometimes the experimental and control groups are created by finding pairs of matching subjects and assigning one to the experimental group and the other to the control group.

© Cengage Learning®

Matching

Another way to achieve comparability between the experimental and control groups is through **matching**. This process is similar to the quota-sampling methods discussed in Chapter 7. If 12 of our subjects are young white men, we might assign 6 of them at random to the experimental group and the other 6 to the control group. If 14 are middle-aged African American women, we might assign 7 to each group. We repeat this process for every relevant grouping of subjects.

The overall matching process could be most efficiently achieved through the creation of a quota matrix constructed of all the most relevant characteristics. Figure 8-2 provides a simplified illustration of such a matrix. In this example, the experimenter has decided that the relevant characteristics are race, age, and gender. Ideally, the quota matrix is constructed to result in an even number of subjects in each cell of the matrix. Then, half the subjects in each cell go into the experimental group and half into the control group.

Alternatively, we might recruit more subjects than our experimental design requires. We might then examine many characteristics of the large initial group of subjects. Whenever we discover a pair of quite similar subjects, we might assign one at random to the experimental group and the other to the control group. Potential subjects who are unlike anyone else in the initial group might be left out of the experiment altogether.

Whatever method we employ, the desired result is the same. The overall average description of the experimental group should be the same as that of the control group. For example, on average both groups should have about the same ages, the same sex composition, the same racial composition, and so forth. This test of comparability should be used whether the two groups are created through probability sampling or through randomization.

Thus far I've referred to the "relevant" variables without saying clearly what those variables are. Of course, these variables cannot be specified in any definite way, any more than I could specify in Chapter 7 which variables should be used in stratified sampling. Which variables are relevant ultimately depends on the nature and purpose of the experiment. As a general rule, however, the control and experimental groups should be comparable in terms of those variables that are most likely to be related to the dependent variable under study. In a study of prejudice, for example, the two groups should be alike in terms of education, ethnicity, and age, among

matching In connection with experiments, the procedure whereby pairs of subjects are matched on the basis of their similarities on one or more variables, and one member of the pair is assigned to the experimental group and the other to the control group.

other characteristics. In some cases, moreover, we may delay assigning subjects to experimental and control groups until we have initially measured the dependent variable. Thus, for example, we might administer a questionnaire measuring subjects' prejudice and then match the experimental and control groups on this variable to assure ourselves that the two groups exhibit the same overall level of prejudice.

Matching or Randomization?

When assigning subjects to the experimental and control groups, you should be aware of two arguments in favor of randomization over matching. First, you may not be in a position to know in advance which variables will be relevant for the matching process. Second, most of the statistics used to analyze the results of experiments assume randomization. Failure to design your experiment that way, then, makes your later use of those statistics less meaningful.

On the other hand, randomization only makes sense if you have a fairly large pool of subjects, so that the laws of probability sampling apply. With only a few subjects, matching would be a better procedure.

Sometimes researchers can combine matching and randomization. When conducting an experiment on the educational enrichment of young adolescents, for example, J. Milton Yinger and his colleagues (1977) needed to assign a large number of students, aged 13 and 14, to several different experimental and control groups to ensure the comparability of students composing each of the groups. They achieved this goal by the following method.

Beginning with a pool of subjects, the researchers first created strata of students nearly identical to one another in terms of some 15 variables. From each of the strata, students were randomly assigned to the different experimental and control groups. In this fashion, the researchers actually improved on conventional randomization. Essentially, they had used a stratified-sampling procedure (Chapter 7), except that they had employed far more stratification variables than are typically used in, say, survey sampling.

Thus far I've described the classical experiment—the experimental design that best represents the logic of causal analysis in the laboratory. In practice, however, social researchers use a great variety of experimental designs. Let's look at some now.

Variations on Experimental Design

Donald Campbell and Julian Stanley (1963), in a classic book on research design, describe 16 different experimental and quasi-experimental designs. This section summarizes a few of these variations to better show the potential for experimentation in social research.

Preexperimental Research Designs

To begin, Campbell and Stanley discuss three "preexperimental" designs, not to recommend them but because they're frequently used in less-than-professional research. These designs are called preexperimental to indicate that they do not meet the scientific standards of experimental designs, and sometimes they may be used because the conditions for full-fledged experiments are impossible to meet. In the first such design—the *one-shot case study*—the researcher measures a single group of subjects on a dependent variable following the administration of some experimental stimulus. Suppose, for example, that we show the Muslim history film, mentioned earlier, to a group of people and then administer a questionnaire that seems to measure prejudice against Muslims. Suppose further that the answers given to the questionnaire seem to represent a low level of prejudice. We might be tempted to conclude that the film reduced prejudice. Lacking a pretest, however, we can't be sure. Perhaps the questionnaire doesn't really represent a sensitive measure of prejudice, or perhaps the group we're studying was low in prejudice to begin with. In either case, the film might have made no difference, though our experimental results might have misled us into thinking it did.

The second preexperimental design discussed by Campbell and Stanley adds a pretest for the experimental group but lacks a control group. This design—which the authors call the *one-group pretest–posttest design*—suffers from the possibility that some factor other than the independent variable might cause a change between

the pretest and posttest results, such as the assassination of a respected Muslim leader. Thus, although we can see that prejudice has been reduced, we can't be sure that the film is what caused that reduction.

To round out the possibilities for preexperimental designs, Campbell and Stanley point out that some research is based on experimental and control groups but has no pretests. They call this design the *static-group comparison*. For example, we might show the Muslim history film to one group and not to another and then measure prejudice in both groups. If the experimental group had less prejudice at the conclusion of the experiment, we might assume the film was responsible. But unless we had randomized our subjects, we would have no way of knowing that the two groups had the same degree of prejudice initially; perhaps the experimental group started out with less.

Figure 8-3 graphically illustrates these three preexperimental research designs by using a different research question: Does exercise cause weight reduction? To make the several designs
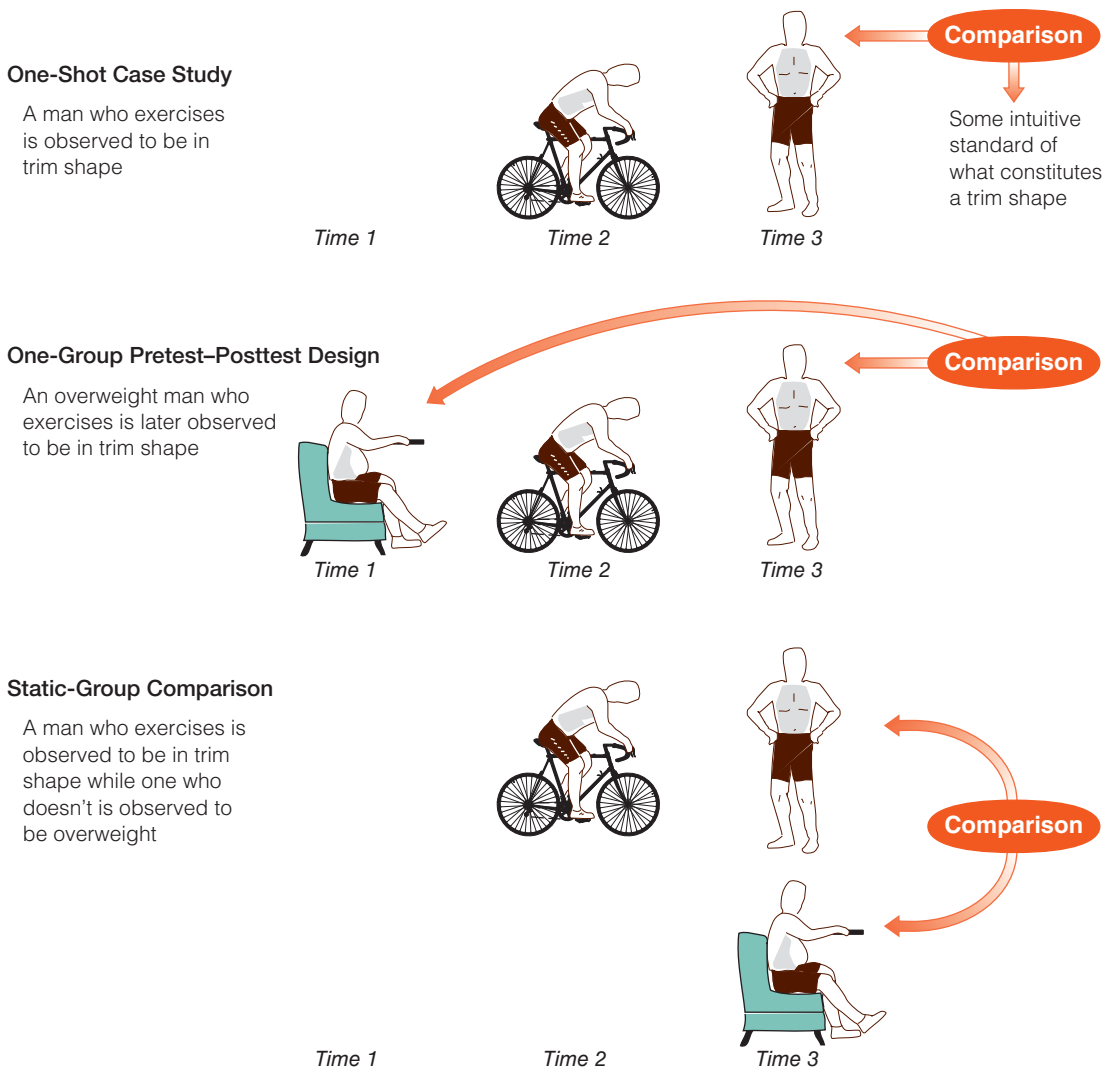
One-Shot Case Study

A man who exercises is observed to be in trim shape

Time 1 *Time 2* *Time 3*

Comparison

Some intuitive standard of what constitutes a trim shape

One-Group Pretest–Posttest Design

An overweight man who exercises is later observed to be in trim shape

Time 1 *Time 2* *Time 3*

Comparison

Static-Group Comparison

A man who exercises is observed to be in trim shape while one who doesn't is observed to be overweight

Time 1 *Time 2* *Time 3*

Comparison

FIGURE 8-3

Three Preexperimental Research Designs. These preexperimental designs anticipate the logic of true experiments but leave themselves open to errors of interpretation. Can you see the errors that might be made in each of these designs? The various risks are solved by the addition of control groups, pretesting, and posttesting.

© Cengage Learning®

clearer, the figure shows individuals rather than groups, but the same logic pertains to group comparisons. Let's review the three preexperimental designs in this new example.

The one-shot case study represents a common form of logical reasoning in everyday life. Asked whether exercise causes weight reduction, we may bring to mind an example that would seem to support the proposition: someone who exercises and is thin. There are problems with this reasoning, however. Perhaps the person was thin long before beginning to exercise. Or perhaps he became thin for some other reason, like eating less or getting sick. The observations shown in the diagram do not guard against these other possibilities. Moreover, the observation that the man in the diagram is in trim shape depends on our intuitive idea of what constitutes trim and overweight body shapes. All told, this is very weak evidence for testing the relationship between exercise and weight loss.

The one-group pretest–posttest design offers somewhat better evidence that exercise produces weight loss. Specifically, we've ruled out the possibility that the man was thin before beginning to exercise. However, we still have no assurance that his exercising is what caused him to lose weight.

Finally, the static-group comparison eliminates the problem of our questionable definition of what constitutes trim or overweight body shapes. In this case, we can compare the shapes of the man who exercises and the one who does not. This design, however, reopens the possibility that the man who exercises was thin to begin with. Notice, this is the same as the posttest-only design, mentioned earlier.

Validity Issues in Experimental Research

At this point I want to present, in a more systematic way, the factors that affect the validity of experimental research. First we'll look at what Campbell and Stanley call the sources of internal invalidity, reviewed and expanded in a follow-up

book by Thomas Cook and Donald Campbell (1979). Then we'll consider the problem of generalizing experimental results to the "real" world, referred to as external invalidity. Having examined these, we'll be in a position to appreciate the advantages of some of the more sophisticated experimental and quasi-experimental designs social science researchers sometimes use.

Sources of Internal Invalidity

The problem of **internal invalidity** refers to the possibility that the conclusions drawn from experimental results may not accurately reflect what has gone on in the experiment itself. The threat of internal invalidity is present whenever anything other than the experimental stimulus can affect the dependent variable.

Donald Campbell and Julian Stanley (1963: 5–6) and Thomas Cook and Donald Campbell (1979: 51–55) point to several sources of internal invalidity. I will touch on eight of them here to illustrate this concern:

1. *History.* During the course of the experiment, historical events may occur that confound the experimental results. The assassination of a Muslim leader during the course of an experiment on reducing anti–Muslim prejudice is one example.

2. *Maturation.* People are continually growing and changing, and such changes affect the results of the experiment. In a long-term experiment, the fact that the subjects grow older (and wiser?) can have an effect. In shorter experiments, they can grow tired, sleepy, bored, or hungry—or change in other ways that affect their behavior in the experiment.

3. *Testing.* Often the process of testing and retesting influences people's behavior, thereby confounding the experimental results. Suppose we administer a questionnaire to a group as a way of measuring their prejudice. Then we administer an experimental stimulus and remeasure their prejudice. As we saw earlier, by the time we conduct the posttest, the subjects will probably have become more sensitive to the issue of prejudice and will be more thoughtful in their answers. In fact, they may have figured out that we're trying to find out how prejudiced they are, and,

internal invalidity Refers to the possibility that the conclusions drawn from experimental results may not accurately reflect what went on in the experiment itself.

because few people want to appear prejudiced, they may give answers that they think the researchers are seeking or that will make themselves "look good."

4. *Instrumentation.* The process of measurement in pretesting and posttesting brings in some of the issues of conceptualization and operationalization discussed earlier in the book. For example, if we use different measures of the dependent variable (say, different questionnaires about prejudice), how can we be sure they're comparable? Perhaps prejudice will seem to decrease simply because the pretest measure was more sensitive than the posttest measure. Or if the measurements are being made by the experimenters, their standards or abilities may change over the course of the experiment.

5. *Statistical regression.* Sometimes it's appropriate to conduct experiments on subjects who start out with extreme scores on the dependent variable. If you were testing a new method for teaching math to hard-core failures in math, you would want to conduct your experiment on people who previously have done extremely poorly in math. But consider for a minute what's likely to happen to the math achievement of such people over time without any experimental interference. They're starting out so low that they can only stay at the bottom or improve: They can't get worse. Even without any experimental stimulus, then, the group as a whole is likely to show some improvement over time. Referring to a *regression to the mean,* statisticians often point out that extremely tall people as a group are likely to have children shorter than themselves, and extremely short people as a group are likely to have children taller than themselves. There is a danger, then, that changes occurring by virtue of subjects starting out in extreme positions will be attributed erroneously to the effects of the experimental stimulus.

6. *Selection biases.* We discussed selection bias earlier when we examined different ways of selecting subjects for experiments and assigning them to experimental and control groups. Comparisons don't have any meaning unless the groups are comparable at the start of an experiment.

7. *Experimental mortality.* We discussed selection bias earlier when we examined different ways of selecting subjects for experiments and assigning them to experimental and control groups. Comparisons have no meaning unless the groups are comparable at the start of an experiment.

8. *Demoralization.* On the other hand, feelings of deprivation within the control group may result in some giving up. In educational experiments, control-group subjects may feel the experimental group is being treated better and they may become demoralized, stop studying, act up, or get angry.

These, then, are some of the sources of internal invalidity in experiments, as cited by Campbell, Stanley, and Cook. Aware of these pitfalls, experimenters have devised designs aimed at managing them. The classical experiment, coupled with proper subject selection and assignment, addresses each of these problems. Let's look again at that study design, presented in Figure 8-4, as it applies to our hypothetical study of prejudice.

If we use the experimental design shown in Figure 8-4, we should expect two findings from our Muslim history film experiment. For the experimental group, the level of prejudice measured in their posttest should be less than was found in their pretest. In addition, when the two posttests are compared, less prejudice should be found in the experimental group than in the control group.

This design also guards against the problem of history, in that anything occurring outside the experiment that might affect the experimental group should also affect the control group. Consequently, the two posttest results should still differ. The same comparison guards against problems of maturation as long as the subjects have been randomly assigned to the two groups. Testing and instrumentation can't be problems, because both the experimental and control groups are subject to the same tests and experimenter effects. If the subjects have been assigned to the two groups randomly, statistical regression should affect both equally, even if people with extreme scores on prejudice (or whatever the dependent variable is) are being studied. Selection bias is ruled out by the random assignment of subjects. Experimental mortality is more complicated to handle, but the data provided in this
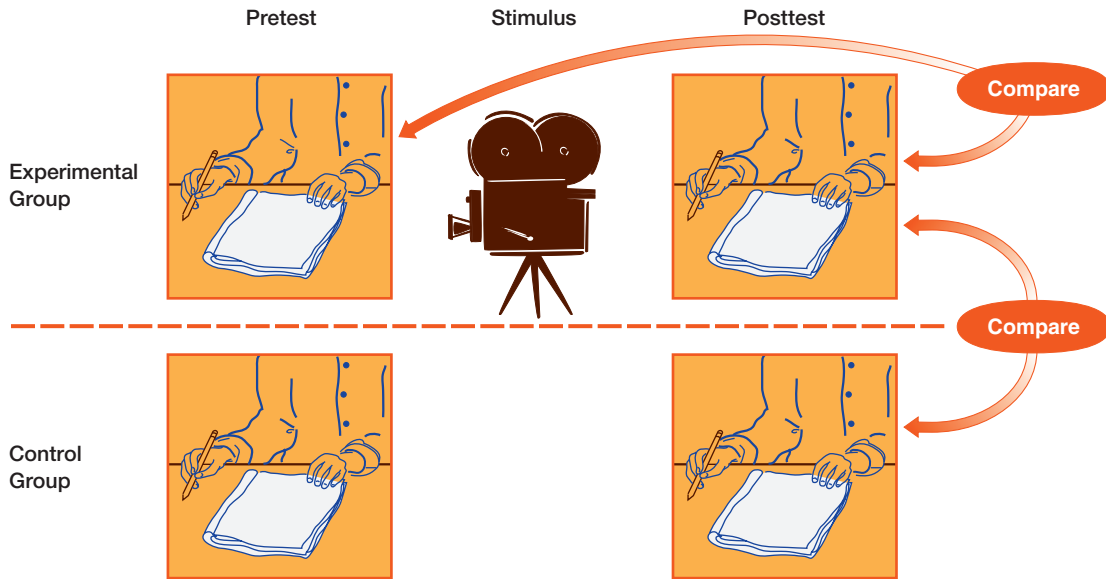
FIGURE 8-4

The Classical Experiment: Using a Muslim History Film to Reduce Prejudice. This diagram illustrates the basic structure of the classical experiment as a vehicle for testing the impact of a film on prejudice. Notice how the control group, the pretesting, and the posttesting function.

© Cengage Learning®

study design offer several ways to deal with it. Pretest measurements would let us discover any differences in the dropouts of the experimental and control groups. Slight modifications to the design—administering a placebo (such as a film having nothing to do with Muslims) to the control group, for example—can make the problem even easier to manage. Finally, demoralization can be watched for and taken into account in evaluating the results of the experiment.

Sources of External Invalidity

Internal invalidity accounts for only some of the complications faced by experimenters. In addition, there are problems of what Campbell and Stanley call **external invalidity**, which relates to the generalizability of experimental findings to the "real" world. Even if the results of an experiment provide an accurate gauge of what happened during that experiment, do they really tell us anything about life in the wilds of society?

Campbell and Stanley describe four forms of this problem; I'll present one of them as an illustration. The generalizability of experimental findings is jeopardized, as the authors point out, if there's an interaction between the testing situation and the experimental stimulus (1963: 18). Here's an example of what they mean.

Staying with the study of prejudice and the Muslim history film, let's suppose that our experimental group—in the classical experiment—has less prejudice in its posttest than in its pretest and that its posttest shows less prejudice than that of the control group. We can be confident that the film actually reduced prejudice among our experimental subjects. But would it have the same effect if the film were shown in theaters or on television? We can't be sure, because the film might be effective only when people have been sensitized to the issue of prejudice, as the subjects may have been in taking the pretest. This is an example of interaction between the testing and the stimulus. The classical experimental design cannot control for that possibility. Fortunately, experimenters have devised other designs that can.

The *Solomon four-group design* (D. Campbell and Stanley 1963: 24–25) addresses the problem of testing interaction with the stimulus. As the name suggests, it involves four groups of subjects, assigned randomly from a pool. Figure 8-5 presents this design graphically.

external invalidity Refers to the possibility that conclusions drawn from experimental results may not be generalizable to the "real" world.
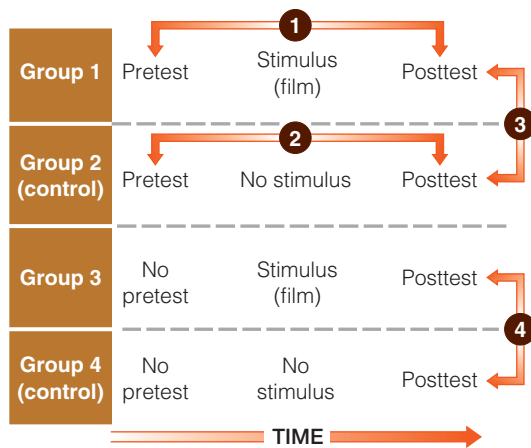
Expected Findings

1 In Group 1, posttest prejudice should be less than pretest prejudice.

2 In Group 2, prejudice should be the same in the pretest and the posttest.

3 The Group 1 posttest should show less prejudice than the Group 2 posttest does.

4 The Group 3 posttest should show less prejudice than the Group 4 posttest does.

FIGURE 8-5

The Solomon Four-Group Design. The classical experiment runs the risk that pretesting will have an effect on subjects, so the Solomon four-group design adds experimental and control groups that skip the pretest. Thus, it combines the classical experiment and the after-only design (with no pretest).

© Cengage Learning®

Notice that Groups 1 and 2 in Figure 8-5 compose the classical experiment, with Group 2 being the control group. Group 3 is administered the experimental stimulus without a pretest, and Group 4 is only posttested. This experimental design permits four meaningful comparisons, which are described in the figure. If the Muslim history film really reduces prejudice—unaccounted for by the problem of internal validity and unaccounted for by an interaction between the testing and the stimulus—we should expect four findings:

1. In Group 1, posttest prejudice should be less than pretest prejudice.

2. In Group 2, prejudice should be the same in the pretest and the posttest.

3. The Group 1 posttest should show less prejudice than the Group 2 posttest.

4. The Group 3 posttest should show less prejudice than the Group 4 posttest.

Notice that Finding 4 rules out any interaction between the testing and the stimulus. And remember that these comparisons are meaningful only if subjects have been assigned randomly to the different groups, thereby providing groups of equal prejudice initially, even though their preexperimental prejudice is measured only in Groups 1 and 2.

There is a side benefit to this research design, as the authors point out. Not only does the Solomon four-group design rule out interactions between testing and the stimulus, it also provides data for comparisons that will reveal how much of this interaction has occurred in a classical experiment. This knowledge allows a researcher to review and evaluate the value of any prior research that used the simpler design.

The last experimental design I'll mention here is what Campbell and Stanley (1963: 25–26) call the *posttest-only control-group design;* it consists of the second half—Groups 3 and 4—of the Solomon design. As the authors argue persuasively, with proper randomization, only Groups 3 and 4 are needed for a true experiment that controls for the problems of internal invalidity as well as for the interaction between testing and stimulus. With randomized assignment to experimental and control groups (which distinguishes this design from the static-group comparison discussed earlier), the subjects will be initially comparable on the dependent variable—comparable enough to satisfy the conventional statistical tests used to evaluate the results—so it's not necessary to measure them. Indeed, Campbell and Stanley suggest that the only justification for pretesting in this situation is tradition. Experimenters have simply grown accustomed to pretesting and feel more secure with research designs that include it. Be clear, however, that this point applies only to experiments in which subjects have been assigned to experimental and control groups randomly, because that's what justifies the assumption that the groups are equivalent without having been measured to find out.

This discussion has introduced the intricacies of experimental design, its problems, and some solutions. There are, of course, a great many other experimental designs in use. Some involve more than one stimulus and combinations of stimuli. Others involve several tests of the dependent variable over time and the administration of the stimulus at different times for different

groups. If you're interested in pursuing this topic, you might want to look at the Campbell and Stanley book.

An Illustration of Experimentation

Experiments have been used to study a wide variety of topics in the social sciences. Some experiments have been conducted within laboratory situations; others occur out in the "real world" and are referred to as *field experiments.* The following discussion provides a glimpse of both. We'll begin with an example of a field experiment.

In George Bernard Shaw's well-loved play *Pygmalion*—the basis of the long-running Broadway musical *My Fair Lady*—Eliza Doolittle speaks of the powers others have in determining our social identity. Here's how she distinguishes the way she's treated by her tutor, Professor Higgins, and by Higgins's friend, Colonel Pickering:

> You see, really and truly, apart from the things anyone can pick up (the dressing and the proper way of speaking, and so on), the difference between a lady and a flower girl is not how she behaves, but how she's treated. I shall always be a flower girl to Professor Higgins, because he always treats me as a flower girl, and always will, but I know I can be a lady to you, because you always treat me as a lady, and always will.
>
> *(Act V)*

The sentiment Eliza expresses here is basic social science, addressed more formally by sociologists such as Charles Horton Cooley (the "looking-glass self") and George Herbert Mead ("the generalized other"). The basic point is that who we think we are—our self-concept—and how we behave are largely a function of how others see and treat us. Related to this, the way others perceive us is largely conditioned by expectations they have in advance. If they've been told we're stupid, for example, they're likely to see us that way—and we may come to see ourselves that way and, in fact, actually act stupidly. "Labeling theory" addresses the phenomenon of people acting in accord with the ways that others perceive and label them. These theories have served

as the premise for numerous movies, such as the 1983 film *Trading Places,* in which Eddie Murphy and Dan Aykroyd play a derelict converted into a stockbroker and vice versa.

The tendency to see in others what we've been led to expect takes its name from Shaw's play. Called the "Pygmalion effect," it's nicely suited to controlled experiments. In one of the best-known experimental investigations of the Pygmalion effect, Robert Rosenthal and Lenore Jacobson (1968) administered what they called the "Harvard Test of Inflected Acquisition" to students in a West Coast school. Subsequently, they met with the students' teachers to present the results of the test. In particular, Rosenthal and Jacobson identified certain students as very likely to exhibit a sudden spurt in academic abilities during the coming year, based on the results of the test.

When IQ test scores were compared later, the researchers' predictions proved accurate. The students identified as "spurters" far exceeded their classmates during the following year, suggesting that the predictive test was a powerful one. In fact, the test was a hoax! The researchers had made their predictions randomly among both good and poor students. What they told the teachers did not really reflect students' test scores at all. The progress made by the "spurters" was simply a result of the teachers expecting the improvement and paying more attention to those students, encouraging them, and rewarding them for achievements. (Notice the similarity between this situation and the Hawthorne effect discussed earlier in this chapter.)

The Rosenthal–Jacobson study attracted a great deal of popular as well as scientific attention. Subsequent experiments have focused on specific aspects of what has become known as the *attribution process,* or the *expectations communication model.* This research, largely conducted by psychologists, parallels research primarily by sociologists, which takes a slightly different focus and is often gathered under the label *expectations-states theory.* Psychological studies focus on situations in which the expectations of a dominant individual affect the performance of subordinates—as in the case of a teacher and students, or a boss and employees. The sociological research has tended to focus more on the role of expectations among equals in small,

task-oriented groups. In a jury, for example, how do jurors initially evaluate each other, and how do those initial assessments affect their later interactions? (You can learn more about this phenomenon, including attempts to find practical applications, by searching the web for "Pygmalion effect.")

Here's an example of an experiment conducted to examine the way our perceptions of our abilities and the abilities of others affect our willingness to accept the other person's ideas. Martha Foschi, G. Keith Warriner, and Stephen Hart (1985) were particularly interested in the role "standards" play in that respect:

> In general terms, by "standards" we mean how well or how poorly a person has to perform in order for an ability to be attributed or denied him/her. In our view, standards are a key variable affecting how evaluations are processed and what expectations result. For example, depending on the standards used, the same level of success may be interpreted as a major accomplishment or dismissed as unimportant.
>
> *(1985: 108–9)*

To begin examining the role of standards, the researchers designed an experiment involving four experimental groups and a control. Subjects were told that the experiment involved something called "pattern recognition ability," defined as an innate ability some people had and others did not. The researchers said subjects would be working in pairs on pattern recognition problems.

In fact, of course, there's no such thing as pattern recognition ability. The object of the experiment was to determine how information about this supposed ability affected subjects' subsequent behavior.

The first stage of the experiment was to "test" each subject's pattern recognition abilities. If you had been a subject in the experiment, you would have been shown a geometric pattern for eight seconds, followed by two more patterns, each of which was similar to but not the same as the first one. Your task would be to choose which of the subsequent set had a pattern closest to the first one you saw. You would be asked to do this 20 times, and a computer would print out your "score." Half the subjects would be told that

they had gotten 14 correct; the other half would be told that they had gotten only 6 correct—regardless of which patterns they matched with which. Depending on the luck of the draw, you would think you had done either quite well or quite badly. Notice, however, that you wouldn't really have any standard for judging your performance—maybe getting 4 correct would be considered a great performance.

At the same time you were given your score, however, you would also be given your "partner's score," although both the "partners" and their "scores" would also be computerized fictions. (Subjects were told they would be communicating with their partners via computer terminals but would not be allowed to see each other.) If you were assigned a score of 14, you would be told your partner had a score of 6; if you were assigned 6, you would be told your partner had 14.

This procedure meant that you would enter the teamwork phase of the experiment believing either (1) you had done better than your partner or (2) you had done worse than your partner. This information constituted part of the "standard" you would be operating under in the experiment. In addition, half of each group was told that a score of between 12 and 20 meant the subject *definitely* had pattern recognition ability; the other subjects were told that a score of 14 wasn't really high enough to prove anything definite. Thus, you would emerge from this with one of the following beliefs:

1. You are *definitely better* at pattern recognition than your partner.
2. You are *possibly better* than your partner.
3. You are *possibly worse* than your partner.
4. You are *definitely worse* than your partner.

The control group for this experiment was told nothing about their own abilities or those of their partners. In other words, they had no expectations.

The final step in the experiment was to set the "teams" to work. As before, you and your partner would be given an initial pattern, followed by a comparison pair to choose from. When you entered your choice in this round, however, you would be told what your partner had answered; then you would be asked to choose again. In your

final choice, you could either stick with your original choice or switch. The "partner's" choice was, of course, created by the computer, and as you can guess, there were often disagreements in the teams: 16 out of 20 times, in fact.

The dependent variable in this experiment was the extent to which subjects would switch their choices to match those of their partners. The researchers hypothesized that the *definitely better* group would switch least often, followed by the *possibly better* group, followed by the *control group,* followed by the *possibly worse* group, followed by the *definitely worse* group, who would switch most often.

The number of times subjects in the five groups switched their answers follows. Realize that each had 16 opportunities to do so. These data indicate that each of the researchers' expectations was correct—with the exception of the comparison between the *possibly worse* and *definitely worse* groups. Although the latter group was in fact the more likely to switch, the difference was too small to be taken as a confirmation of the hypothesis. (Chapter 16 will discuss the statistical tests that let researchers make decisions like this.)

Group	Mean Number of Switches
Definitely better	5.05
Possibly better	6.23
Control group	7.95
Possibly worse	9.23
Definitely worse	9.28

In more-detailed analyses, it was found that the same basic pattern held for both men and women, though it was somewhat clearer for women than for men. Here are the actual data:

	Mean Number of Switches	
	Women	Men
Definitely better	4.50	5.66
Possibly better	6.34	6.10
Control group	7.68	8.34
Possibly worse	9.36	9.09
Definitely worse	10.00	8.70

Because specific research efforts like this one sometimes seem extremely focused in their scope, you might wonder about their relevance to anything. As part of a larger research effort, however, studies like this one add concrete pieces to our understanding of more-general social processes.

It's worth taking a minute to consider some of the life situations where "expectation states" might have very real and important consequences. I've mentioned the case of jury deliberations. How about all forms of prejudice and discrimination? Or, consider how expectation states figure into job interviews or meeting your heartthrob's parents. If you think about it, you'll undoubtedly see other situations where these laboratory concepts apply in real life.

Alternative Experimental Settings

Although we tend to equate the terms *experiment* and *laboratory experiment,* many important social science experiments occur outside controlled settings, as we've seen in our example of the Rosenthal–Jacobson study of the Pygmalion effect. Two other special circumstances deserve mention here: web-based experiments and "natural" experiments.

Here's a different kind of social science experiment. Shelley J. Correll, Stephen Benard, and In Paik (2007) were interested in learning whether race, gender, and/or parenthood might produce discrimination in hiring. Specifically, they wanted to find out if there was a "Motherhood penalty." These researchers decided to explore this topic with an experiment using college undergraduates. The student-subjects chosen for the study were told that a new communications company was looking for someone to manage the marketing department of their East Coast office.

They heard that the communications company was interested in receiving feedback from younger adults since young people are heavy consumers of communications technology. To further increase their task orientation, participants were told that their input would be incorporated with the other information the company collects on applicants and would impact actual hiring decisions.

(2007: 1311)

The researchers had created a number of resumes describing fictitious candidates for the manager's position. Initially, the resumes had no indication of race, sex, or parenthood, and a group of subjects was asked to evaluate the quality of the candidates. The initial evaluations showed the resumes to be equivalent in apparent quality.

Then, in the main experiment, the resumes were augmented with additional information. Gender became apparent when names were added to the resumes. Moreover, the use of typically African American names (e.g., Latoya and Ebony for women; Tyrone and Jamal for men) or typically white names (e.g., Allison and Sarah for women; Brad and Matthew for men) allowed subjects to guess the candidates' races. Finally, listing participation in a Parent–Teacher Association or listing names of children identified some candidates as parents. Over the course of the experiment, these different status indicators were added to the same resumes. Thus a particular resume might appear as a black mother, a white non-mother, a white father, and so forth. Of course, no student-subject would evaluate the same resume with different status indicators.

Finally, the experimental subjects were given sets of resumes to evaluate in a number of ways. For example, they were asked how competent they felt the candidates were and how committed they seemed. They were asked to suggest a salary that might be offered a given candidate and to predict how likely it was that the candidate would eventually be promoted within the organization. They were even asked to indicate how many days the candidate should be allowed to miss work or come late before being fired.

Since each of the resumes was evaluated with different status indicators attached, it was possible for the experimenters to determine whether those statuses made a difference. Specifically, they could test for the existence of a Motherhood penalty. And they found it. Among other things:

- Mothers were judged less competent and less committed than non-mothers.
- Students offered the mothers lower salaries than the non-mothers and would allow them fewer missed or late days on the job.
- They felt the mothers were less likely to be promoted than the non-mothers.

- And they were almost twice as likely to recommend hiring the non-mothers.

Rounding out the analysis of gender and parenthood, the researchers found that, while the differences were smaller for men than for women, fathers were rated *higher* than non-fathers. This was just the opposite pattern as had been found among women candidates.

The Motherhood penalty was found among both white and African American candidates. Moreover, it did not matter what the gender of the subject evaluators were. Both women and men rated mothers lower than non-mothers.

Factorial Designs

Up to now, I have discussed the experimental variable as singular: We try to limit the variation between experimental and control group to one variable. While this logic is basic to the experimental model, **factorial designs** expand that model to encompass more than one experimental variable. Let's say we are interested in what brings consumers to hunger for Green Healthy Treats (GHT). Are they more moved by environmental or health issues?

Let's suppose we create TV spots that (1) emphasize the environmental value of the way GHT is produced and (2) and how healthy it is for you. We produce two ads, let's call them E and H to reflect Environmental and Health emphases. Now, instead of having one experimental group, we have three:

> E only
> H only
> E & H both

Now we can compare the desire for GHT among those who were shown the Environmental ad only (E), the Health ad only (H), and both ads (E & H). This design enables us to determine whether (a) the Environmental ad makes a difference, regardless of whether viewers saw the Health ad; (b) the Environmental ad makes a difference regardless of whether they saw the Environmental ad; (c) these two ads have independent, cumulative support for using GHT; or (d) neither ad makes a difference.

factorial design An experimental design using more than one experimental variable.

Web-Based Experiments

Increasingly, researchers are using the Internet as a vehicle for conducting social science experiments. Because representative samples are not essential in most experiments, researchers can often use volunteers who respond to invitations online. One site you might visit to get a better idea of this form of experimentation is Online Social Psychology Studies. This website offers hot links to numerous professional and student research projects on such topics as "interpersonal relations," "beliefs and attitudes," and "personality and individual differences." In addition, the site offers some resources for conducting web experiments.

"Natural" Experiments

Important social science experiments can occur in the course of normal social events, outside controlled settings. Sometimes nature designs and executes experiments that we can observe and analyze; sometimes social and political decision makers serve this natural function.

Imagine, for example, that a hurricane has struck a particular town. Some residents of the town suffer severe financial damages, and others escape relatively lightly. What, we might ask, are the behavioral consequences of suffering a natural disaster? Are those who suffer most more likely to take precautions against future disasters than are those who suffer least? To answer these questions, we might interview residents of the town some time after the hurricane. We might question them regarding the precautions they had taken before the hurricane and those they're currently taking toward future preparedness. We could then compare the precautionary actions of the people who suffered a great deal from the hurricane with those taken by citizens who suffered relatively little. In this fashion, we might take advantage of a natural experiment, which we could not have arranged even if we'd been perversely willing to do so.

Because the researcher must, for the most part, take things as they occur, natural experiments raise many of the validity problems discussed earlier. Thus, when Stanislav Kasl, Rupert Chisolm, and Brenda Eskenazi (1981) chose to study the impact that the Three Mile Island (TMI) nuclear accident in Pennsylvania had on

plant workers, they had to be especially careful while devising the study design:

> Disaster research is necessarily opportunistic, quasi-experimental, and after-the-fact. In the terminology of Campbell and Stanley's classical analysis of research designs, our study falls into the "static-group comparison" category, considered one of the weak research designs. However, the weaknesses are potential and their actual presence depends on the unique circumstances of each study.
>
> *(1981: 474)*

The foundation of this study was a survey of the people who had been working at Three Mile Island on March 28, 1979, when the cooling system failed in the number 2 reactor and began melting the uranium core. The survey was conducted five to six months after the accident. Among other things, the survey questionnaire measured workers' attitudes toward working at nuclear power plants. If they had measured only the TMI workers' attitudes after the accident, the researchers would have had no idea whether attitudes had changed as a consequence of the accident. But they improved their study design by selecting another, nearby—seemingly comparable—nuclear power plant (abbreviated as PB) and surveyed workers there as a control group: hence their reference to a static-group comparison.

Even with an experimental and a control group, the authors were wary of potential problems in their design. In particular, their design was based on the idea that the two sets of workers were equivalent to each other, except for the single fact of the accident. The researchers could have assumed this if they had been able to assign workers to the two plants randomly, but of course that was not the case. Instead, they needed to compare characteristics of the two groups and infer whether or not they were equivalent. Ultimately, the researchers concluded that the two sets of workers were very much alike, and the plant the employees worked at was merely a function of where they lived.

Even granting that the two sets of workers were equivalent, the researchers faced another problem of comparability. They could not contact all the workers who had been employed at TMI at the time of the accident. The researchers discussed the problem as follows:

One special attrition problem in this study was the possibility that some of the no-contact nonrespondents among the TMI subjects, but not PB subjects, had permanently left the area because of the accident. This biased attrition would, most likely, attenuate the estimated extent of the impact. Using the evidence of disconnected or "not in service" telephone numbers, we estimate this bias to be negligible (1 percent).

(Kasl, Chisolm, and Eskenazi 1981: 475)

The TMI example points to both the special problems involved in natural experiments and the possibility for taking those problems into account. Social research generally requires ingenuity and insight, and natural experiments are certainly no exception. Earlier in this chapter, we used a hypothetical example of studying whether an ethnic history film reduced prejudice. Sandra Ball-Rokeach, Joel Grube, and Milton Rokeach (1981) were able to address that topic in real life through a natural experiment. In 1977, the television dramatization of Alex Haley's *Roots,* a historical saga about African Americans, was presented by ABC on eight consecutive nights. It garnered the largest audiences in television history up to that time. Ball-Rokeach and her colleagues wanted to know whether *Roots* changed white Americans' attitudes toward African Americans. Their opportunity arose in 1979, when a sequel—*Roots: The Next Generation*—was televised. Although it would have been nice (from a researcher's point of view) to assign random samples of Americans either to watch or not to watch the show, that wasn't possible. Instead, the researchers selected four samples in Washington State and mailed questionnaires that measured attitudes toward African Americans. Following the last episode of the show, respondents were called and asked how many, if any, episodes they had watched. Subsequently, questionnaires were sent to respondents, remeasuring their attitudes toward African Americans.

By comparing attitudes before and after for both those who watched the show and those who didn't, the researchers reached several conclusions. For example, they found that people with already egalitarian attitudes were much more likely to watch the show than were those who were more prejudiced toward African Americans: a self-selection phenomenon.

Comparing the before and after attitudes of those who watched the show, moreover, suggested the show itself had little or no effect. Those who watched it were no more egalitarian afterward than they had been before.

This example anticipates the subject of Chapter 12, evaluation research, which can be seen as a special type of natural experiment. As you'll see, evaluation research involves taking the logic of experimentation into the field to observe and evaluate the effects of stimuli in real life. Because this is an increasingly important form of social research, an entire chapter is devoted to it.

Strengths and Weaknesses of the Experimental Method

Experiments are the primary tool for studying causal relationships. However, like all research methods, experiments have both strengths and weaknesses.

The chief advantage of a controlled experiment lies in the isolation of the experimental variable's impact over time. This is seen most clearly in terms of the basic experimental model. A group of experimental subjects are found, at the outset of the experiment, to have a certain characteristic; following the administration of an experimental stimulus, they are found to have a different characteristic. To the extent that subjects have experienced no other stimuli, we may conclude that the change of characteristics is attributable to the experimental stimulus.

Further, because individual experiments are often rather limited in scope, requiring relatively little time and money and relatively few subjects, we often can replicate a given experiment several times using several different groups of subjects. (This isn't always the case, of course, but it's usually easier to repeat experiments than, say, surveys.) As in all other forms of scientific research, replication of research findings strengthens our confidence in the validity and generalizability of those findings.

The greatest weakness of laboratory experiments lies in their artificiality. Social processes that occur in a laboratory setting might not necessarily occur in natural social settings. For example, a Muslim history film might genuinely reduce prejudice among a group of experimental

subjects. This would not necessarily mean, however, that the same film shown in neighborhood movie theaters throughout the country would reduce prejudice among the general public. Artificiality is not as much of a problem, of course, for natural experiments as for those conducted in the laboratory.

In discussing several of the sources of internal and external invalidity mentioned by Campbell, Stanley, and Cook, we saw that we can create experimental designs that logically control such problems. This possibility points to one of the great advantages of experiments: They lend themselves to a logical rigor that is often much more difficult to achieve in other modes of observation.

Ethics and Experiments

As you've probably realized by now, researchers must consider many important ethical issues in conducting social science experiments. I'll mention only two here.

First, experiments almost always involve deception. In most cases, explaining the purpose of the experiment to subjects would probably cause them to behave differently—trying to look less prejudiced, for example. It's important, therefore, to determine (1) whether a particular deception is essential to the experiment and (2) whether the value of what may be learned from the experiment justifies the ethical violation.

Second, experiments are typically intrusive. Subjects often are placed in unusual situations and asked to undergo unusual experiences. Even when the subjects are not physically injured (don't do that, by the way), there is always the possibility that they could be psychologically damaged, as some of the previous examples in this chapter have illustrated. As with the matter of deception, you'll find yourself balancing the potential value of the research against the potential damage to subjects.

MAIN POINTS

Introduction

- In experiments, social researchers typically select a group of subjects, do something to them, and observe the effect of what was done.

Topics Appropriate for Experiments

- Experiments are an excellent vehicle for the controlled testing of causal processes.

The Classical Experiment

- The classical experiment tests the effect of an experimental stimulus (the independent variable) on a dependent variable through the pretesting and posttesting of experimental and control groups.
- It is generally less important that a group of experimental subjects be representative of some larger population than that experimental and control groups be similar to each other.
- A double-blind experiment guards against experimenter bias, because neither the experimenter nor the subject knows which subjects are in the control group(s) and which are in the experimental group(s).

Selecting Subjects

- Probability sampling, randomization, and matching are all methods of achieving comparability in the experimental and control groups. Randomization is the generally preferred method. In some designs, it can be combined with matching.

Variations on Experimental Design

- Campbell and Stanley describe three forms of preexperiments: the one-shot case study, the one-group pretest–posttest design, and the static-group comparison. None of these designs features all the controls available in a true experiment.
- Campbell and Stanley list, among others, eight sources of internal invalidity in experimental design. The classical experiment with random assignment of subjects guards against each of these problems.
- Experiments also face problems of external invalidity: Experimental findings may not reflect real life.
- The interaction of testing and stimulus is an example of external invalidity that the classical experiment does not guard against.
- The Solomon four-group design and other variations on the classical experiment can safeguard against external invalidity.
- Campbell and Stanley suggest that, given proper randomization in the assignment of subjects to the experimental and control groups, there is no need for pretesting in experiments.

An Illustration of Experimentation

- Experiments on "expectation states" demonstrate experimental designs and show how experiments can prove relevant to real-world concerns.

Alternative Experimental Settings

- More and more, researchers are using the Internet for conducting experiments.
- Natural experiments often occur in the course of social life in the real world, and social researchers can implement them in somewhat the same way they would design and conduct laboratory experiments.

Strengths and Weaknesses
of the Experimental Method

- Like all research methods, experiments have strengths and weaknesses. Their primary weakness is artificiality: What happens in an experiment may not reflect what happens in the outside world. Strengths include the isolation of the independent variable, which permits causal inferences; the relative ease of replication; and scientific rigor.

Ethics and Experiments

- Experiments typically involve deceiving subjects.
- By their intrusive nature, experiments open the possibility of inadvertently causing damage to subjects.

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.

control group	internal invalidity
double-blind experiment	matching
experimental group	posttesting
external invalidity	pretesting
factorial design	randomization

PROPOSING SOCIAL RESEARCH: EXPERIMENTS

In the next series of exercises, we'll focus on specific data-collection techniques, beginning with experiments here. If you're doing these exercises as part of an assignment in the course, your instructor will tell you whether you should skip those chapters dealing with methods you won't use. If you're doing these exercises on your own, to improve your understanding of the topics in the book, you can temporarily modify your proposed data-collection method and explore how you would research your topic using the method at hand—in this case, experimentation.

In the proposal, you'll describe the experimental stimulus and how it will be administered, as well as detailing the experimental and control groups you'll use. You'll also describe the pretesting and posttesting that will be involved in your experiment. What will be the setting for your experiments: a laboratory or more-natural circumstances?

It may be appropriate for you to conduct a double-blind experiment, in which case you should describe how you will accomplish it. You may also need to explore some of the internal and external problems of validity that might complicate your analysis of your results.

Finally, the experimental model is used to test specific hypotheses, so you should detail how you will accomplish that in terms of your study.

REVIEW QUESTIONS AND EXERCISES

1. In the library or on the web, locate a research report of an experiment. Identify the dependent variable and the stimulus.

2. Pick 4 of the 8 sources of internal invalidity discussed in this chapter and make up examples (not discussed in the chapter) to illustrate each.

3. Create a hypothetical experimental design that illustrates one of the problems of external invalidity.

4. Think of a recent natural disaster you've witnessed or read about. Frame a research question that might be studied by treating that disaster as a natural experiment. In two or three paragraphs, outline how the study might be done.

5. In this chapter, we looked briefly at the problem of "placebo effects." On the web, find a study in which the placebo effect figured importantly. Write a brief report on the study, including the source of your information. (*Hint:* You might want to do a search on "placebo.")

Survey Research

CHAPTER OVERVIEW

Researchers have many methods for collecting data through surveys—from mail questionnaires to personal interviews to online surveys conducted over the Internet. Social researchers should know how to select an appropriate method and how to implement it effectively.

Introduction

Surveys are a very old research technique. In the Old Testament, for example, we find the following:

> After the plague the Lord said to Moses and to Eleazar the son of Aaron, the priest, "Take a census of all the congregation of the people of Israel, from twenty old and upward."
>
> *(Numbers 26: 1–2)*

Ancient Egyptian rulers conducted censuses to help them administer their domains. Jesus was born away from home because Joseph and Mary were journeying to Joseph's ancestral home for a Roman census.

A little-known survey was attempted among French workers in 1880. A German political sociologist mailed some 25,000 questionnaires to workers to determine the extent of their exploitation by employers. The rather lengthy questionnaire included items such as these:

> Does your employer or his representative resort to trickery in order to defraud you of a part of your earnings?
>
> If you are paid piece rates, is the quality of the article made a pretext for fraudulent deductions from your wages?

The survey researcher in this case was not George Gallup but Karl Marx ([1880] 1956: 208). Though 25,000 questionnaires were mailed out, there is no record of any being returned.

Today, survey research is a frequently used mode of observation in the social sciences. In a typical survey, the researcher selects a sample of respondents and administers a standardized questionnaire to them. Chapter 7 discussed sampling techniques in detail. This chapter discusses how to prepare a questionnaire and describes the various options for administering it so that respondents answer your questions adequately.

This chapter includes a short discussion of secondary analysis, the analysis of survey data collected by someone else. This use of survey results has become an important aspect of survey research in recent years, and it is especially useful for students and others with scarce research funds.

Let's begin by looking at the kinds of topics that researchers can appropriately study by using survey research.

Topics Appropriate for Survey Research

Surveys may be used for descriptive, explanatory, and exploratory purposes. They are chiefly used in studies that have individual people as the units of analysis. Although this method can be employed for other units of analysis, such as groups or interactions, some individual persons must serve as **respondents** or informants. Thus, we could undertake a survey in which divorces were the unit of analysis, but we would need to administer the survey questionnaire to the participants in the divorces (or to some other respondents).

Survey research is probably the best method available to the social researcher who is interested in collecting original data for describing a population too large to observe directly. Careful probability sampling provides a group of respondents whose characteristics may be taken to reflect those of the larger population, and carefully constructed standardized question-naires provide data in the same form from all respondents.

Surveys are also excellent vehicles for measuring attitudes and orientations in a large population. Public opinion polls—for example, Pew, Gallup, Harris, Roper, and a number of university survey centers—are well-known ex-amples of this use. Indeed, polls have become so prevalent that at times the public seems unsure what to think of them. Pollsters are criticized by those who don't think (or want to believe) that polls are accurate (candidates who are "losing" in

respondent A person who provides data for analysis by responding to a survey questionnaire.

polls often tell voters not to trust the polls). But polls are also criticized for being too accurate—as when exit polls on Election Day are used to predict a winner before the actual voting is complete.

The general attitude toward public opinion research is further complicated by scientifically unsound "surveys" that nonetheless capture people's attention because of the topics they cover and/or their "findings." A good example is the "Hite Reports" on human sexuality. While enjoying considerable attention in the popular press, Shere Hite was roundly criticized by the research community for her data-collection methods. For example, a 1987 Hite report was based on questionnaires completed by women around the country—but which women? Hite reported that she distributed some 100,000 questionnaires through various organizations, and around 4,500 were returned.

Now, 4,500 and 100,000 are large numbers in the context of survey sampling. However, given Hite's research methods, her 4,500 respondents didn't necessarily represent U.S. women any more than the *Literary Digest*'s enormous 1936 sample represented the U.S. electorate when their 2 million sample ballots indicated that Alf Landon would bury FDR in a landslide.

Sometimes, people use the pretense of survey research for quite different purposes. For example, you may have received a telephone call indicating you've been selected for a survey, only to find that the first question was "How would you like to make thousands of dollars a week right in your own home?" Or you may have been told you could win a prize if you could name the president whose picture is on the penny. (Tell them it's Elvis.) Unfortunately, a few unscrupulous telemarketers try to prey on the general cooperation people have given to survey researchers.

By the same token, political parties and charitable organizations have begun conducting phony "surveys." Often under the guise of collecting public opinion about some issue,

---

questionnaire A document containing questions and other types of items designed to solicit information appropriate for analysis. Questionnaires are used primarily in survey research but also in experiments, field research, and other modes of observation.

---

callers ultimately ask respondents for a monetary contribution.

Recent political campaigns have produced another form of bogus survey, the "push poll." Here's what the American Association for Public Opinion Polling has said in condemning this practice (see also Figure 3-1):

> A "push poll" is a telemarketing technique in which telephone calls are used to canvass potential voters, feeding them false or misleading "information" about a candidate under the pretense of taking a poll to see how this "information" affects voter preferences. In fact, the intent is not to measure public opinion but to manipulate it—to "push" voters away from one candidate and toward the opposing candidate. Such polls defame selected candidates by spreading false or misleading information about them. The intent is to disseminate campaign propaganda under the guise of conducting a legitimate public opinion poll.
>
> *(Bednarz 1996)*

In short, the labels "survey" and "poll" are sometimes misused. Done properly, however, survey research can be a useful tool of social inquiry. Designing useful (and trustworthy) survey research begins with formulating good questions. Let's turn to that topic now.

Guidelines for Asking Questions

In social research, variables are often operationalized when researchers ask people questions as a way of getting data for analysis and interpretation. Sometimes the questions are asked by an interviewer; sometimes they are written down and given to respondents for completion. In other cases, several general guidelines can help researchers frame and ask questions that serve as excellent operationalizations of variables while avoiding pitfalls that can result in useless or even misleading information.

Surveys include the use of a **questionnaire** — an instrument specifically designed to elicit information that will be useful for analysis. Although some of the specific points to follow are more appropriate to structured questionnaires than to the more open-ended questionnaires used in qualitative, in-depth interviewing, the underlying

logic is valuable whenever we ask people questions in order to gather data.

Choose Appropriate Question Forms

Let's begin with some of the options available to you in creating questionnaires. These options include using questions or statements and choosing **open-ended** or **closed-ended questions**.

Questions and Statements

Although the term *questionnaire* suggests a collection of questions, an examination of a typical questionnaire will probably reveal as many statements as questions. This is not without reason. Often, the researcher is interested in determining the extent to which respondents hold a particular attitude or perspective. If you can summarize the attitude in a fairly brief statement, you can present that statement and ask respondents whether they agree or disagree with it. As you may remember, Rensis Likert greatly formalized this procedure through the creation of the Likert scale, a format in which respondents are asked to strongly agree, agree, disagree, or strongly disagree, or perhaps strongly approve, approve, and so forth.

Both questions and statements can be used profitably. Using both in a given questionnaire gives you more flexibility in the design of items and can make the questionnaire more interesting as well.

Open-Ended and Closed-Ended Questions

In asking questions, researchers have two options. They can ask open-ended questions, in which case the respondent is asked to provide his or her own answers to the questions. For example, the respondent may be asked, "What do you feel is the most important issue facing the United States today?" and be provided with a space to write in the answer (or be asked to report it verbally to an interviewer). As we'll see in Chapter 10, in-depth, qualitative interviewing relies almost exclusively on open-ended questions. However, they are also used in survey research.

In the case of closed-ended questions, the respondent is asked to select an answer from among a list provided by the researcher. Closed-ended questions are very popular in survey

research because they provide a greater uniformity of responses and are more easily processed than open-ended ones.

Open-ended responses must be coded before they can be processed for computer analysis, as we'll see in Chapter 14. This coding process often requires the researcher to interpret the meaning of responses, opening the possibility of misunderstanding and researcher bias. There is also a danger that some respondents will give answers that are essentially irrelevant to the researcher's intent. Closed-ended responses, on the other hand, can often be transferred directly into a computer format.

The chief shortcoming of closed-ended questions lies in the researcher's structuring of responses. When the relevant answers to a given question are relatively clear, there should be no problem. In other cases, however, the researcher's structuring of responses may overlook some important responses. In asking about "the most important issue facing the United States," for example, his or her checklist of issues might omit certain issues that respondents would have said were important.

The construction of closed-ended questions should be guided by two structural requirements. First, the response categories provided should be exhaustive: They should include all the possible responses that might be expected. Often, researchers ensure this by adding a category such as "Other (Please specify: _____)." Second, the answer categories must be mutually exclusive: The respondent should not feel compelled to select more than one. (In some cases, you may wish to solicit multiple answers, but these may create difficulties in data processing and analysis later on.) To ensure that your categories are mutually exclusive, carefully consider each combination of categories, asking yourself whether a person could reasonably

open-ended questions Questions for which the respondent is asked to provide his or her own answers. In-depth, qualitative interviewing relies almost exclusively on open-ended questions.

closed-ended questions Survey questions in which the respondent is asked to select an answer from among a list provided by the researcher. Popular in survey research because they provide a greater uniformity of responses and are more easily processed than open-ended questions.

choose more than one answer. In addition, it's useful to add an instruction to the question asking the respondent to select the one best answer, but this technique is not a satisfactory substitute for a carefully constructed set of responses.

Make Items Clear

It should go without saying that questionnaire items need to be clear and unambiguous, but the broad proliferation of unclear and ambiguous questions in surveys makes the point worth emphasizing. We can become so deeply involved in the topic under examination that opinions and perspectives are clear to us but not to our respondents—many of whom have paid little or no attention to the topic. Or, if we have only a superficial understanding of the topic, we may fail to specify the intent of a question sufficiently. The question "What do you think about the proposed peace plan?" may evoke in the respondent a counter question: "Which proposed peace plan?" Questionnaire items should be precise so that the respondent knows exactly what the researcher is asking. The possibilities for misunderstanding are endless, and no researcher is immune (Polivka and Rothgeb 1993).

One of the most established research projects in the United States is the Census Bureau's ongoing "Current Population Survey" or CPS, which measures, among other critical data, the nation's unemployment rate. A part of the measurement of employment patterns focuses on a respondent's activities during "last week," by which the Census Bureau means Sunday through Saturday. Studies undertaken to determine the accuracy of the survey found that more than half the respondents took "last week" to include only Monday through Friday. By the same token, whereas the Census Bureau defines "working full-time" as 35 or more hours a week, the same evaluation studies showed that some respondents used the more traditional definition of 40 hours per week. As a consequence, the wording of these questions in the CPS was modified in 1994 to specify the Census Bureau's definitions.

Similarly, the use of the term *Native American* to mean American Indian often produces an overrepresentation of that ethnic group in surveys. Clearly, many respondents understand the term to mean "born in the United States."

Avoid Double-Barreled Questions

Frequently, researchers ask respondents for a single answer to a question that actually has multiple parts. These types of queries are often termed *double-barreled questions* and seem to happen most often when the researcher has personally identified with a complex question. For example, you might ask respondents to agree or disagree with the statement "The United States should abandon its space program and spend the money on domestic programs." Although many people would unequivocally agree with the statement and others would unequivocally disagree, still others would be unable to answer. Some would want to abandon the space program and give the money back to the taxpayers. Others would want to continue the space program but also put more money into domestic programs. These latter respondents could neither agree nor disagree without misleading you.

As a general rule, whenever the word *and* appears in a question or questionnaire statement, check whether you're asking a double-barreled question. See the Tips and Tools box, "Double-Barreled and Beyond," for some imaginative variations on this theme.

Respondents Must Be Competent to Answer

In asking respondents to provide information, you should continually ask yourself whether they can do so reliably. In a study of child rearing, you might ask respondents to report the age at which they first talked back to their parents. Quite aside from the problem of defining *talking back to parents*, it's doubtful that most respondents would remember with any degree of accuracy.

As another example, student-government leaders occasionally ask their constituents to indicate how students' fees ought to be spent. Typically, respondents are asked to indicate the percentage of available funds that should be devoted to a long list of activities. Without a fairly good knowledge of the nature of those activities and the costs involved in them, the respondents cannot provide meaningful answers. Administrative costs, for example, will receive little support although they may be essential to the programs as a whole.

One group of researchers examining teenagers' driving experience insisted on asking an

Tips and Tools

Double-Barreled and Beyond

The "Arab Spring" uprisings of 2011 drew world attention to several countries in the Middle East. One of the more dramatic changes culminated with the overthrow of Libya's Colonel Muammar Gaddafi in August. This was not the first time American concerns were focused on Libya.

Consider this question, asked of U.S. citizens in April 1986, at a time when the country's relationship with Libya was at an especially low point. Some observers suggested that the United States might end up in a shooting war with the small North African nation. The Harris Poll sought to find out what U.S. public opinion was.

	U.S. Will Not Go to War	War Is Probable but Not Inevitable	War Is Inevitable
U.S. will not invade Libya	1	2	3
U.S. will invade Libya but it would be wrong		4	5
U.S. will invade Libya and it would be right		6	7

> If Libya now increases its terrorist acts against the U.S. and we keep inflicting more damage on Libya, then inevitably it will all end in the U.S. going to war and finally invading that country, which would be wrong.

Respondents were given the opportunity of answering "Agree," "Disagree," or "Not sure." Notice the elements contained in the complex statement:

1. Will Libya increase its terrorist acts against the U.S.?
2. Will the U.S. inflict more damage on Libya?
3. Will the U.S. inevitably or otherwise go to war against Libya?
4. Would the U.S. invade Libya?
5. Would that be right or wrong?

These several elements offer the possibility of numerous points of view—far more than the three alternatives offered to the survey respondents. Even if we were to assume hypothetically that Libya would "increase its terrorist attacks" and the United States would "keep inflicting more damage" in return, you might have any one of at least seven distinct expectations about the outcome:

The examination of prognoses about the Libyan situation is not the only example of double-barreled questions sneaking into public opinion research. Here are some questions the Harris Poll asked in an attempt to gauge U.S. public opinion about then Soviet General Secretary Gorbachev:

> He looks like the kind of Russian leader who will recognize that both the Soviets and the Americans can destroy each other with nuclear missiles so it is better to come to verifiable arms control agreements.
>
> He seems to be more modern, enlightened, and attractive, which is a good sign for the peace of the world.
>
> Even though he looks much more modern and attractive, it would be a mistake to think he will be much different from other Russian leaders.

How many elements can you identify in each of the questions? How many possible opinions could people have in each case? What does a simple "agree" or "disagree" really mean in such cases?

Sources: Reported in *World Opinion Update,* October 1985 and May 1986, respectively.

---

open-ended question concerning the number of miles driven since receiving a license, even though consultants argued that few drivers could estimate such information with any accuracy. In response, some teenagers reported driving hundreds of thousands of miles.

Respondents Must Be Willing to Answer

Often, we would like to learn things from people that they are unwilling to share with us. For example, Yanjie Bian indicates that it has often been difficult to get candid answers from people in China.

> [Here] people are generally careful about what they say on nonprivate occasions in order to survive under authoritarianism. During the Cultural Revolution between 1966 and 1976, for example, because of the radical political agenda and political intensity throughout the country, it was almost impossible to use survey techniques to collect valid and reliable data inside China about the Chinese people's life experiences, characteristics, and attitudes towards the Communist regime.
>
> *(1994: 19–20)*

Sometimes, U.S. respondents say they're undecided when, in fact, they have an opinion but

think they're in a minority. Under that condition, they may be reluctant to tell a stranger (the interviewer) what that opinion is. Given this problem, the Gallup Organization, for example, has used a "secret ballot" format, which simulates actual election conditions, in that the "voter" enjoys complete anonymity. In an analysis of the Gallup Poll election data from 1944 to 1988, Andrew Smith and G. F. Bishop (1992) have found that this technique substantially reduced the percentage of respondents who said they were undecided about how they would vote.

This problem of nondisclosure is not limited to survey research, however. Richard Mitchell (1991: 100) faced a similar problem in his field research among U.S. survivalists:

> Survivalists, for example, are ambivalent about concealing their identities and inclinations. They realize that secrecy protects them from the ridicule of a disbelieving majority, but enforced separatism diminishes opportunities for recruitment and information exchange. . . .
>
> "Secretive" survivalists eschew telephones, launder their mail through letter exchanges, use nicknames and aliases, and carefully conceal their addresses from strangers. Yet once I was invited to group meetings, I found them cooperative respondents.

Questions Should Be Relevant

Similarly, questions asked in a questionnaire should be relevant to most respondents. When attitudes are requested on a topic that few respondents have thought about or really care about, the results are not likely to be useful. Of course, because the respondents may express attitudes even though they've never given any thought to the issue, you run the risk of being misled.

This point is illustrated occasionally when researchers ask for responses relating to fictitious people and issues. In one political poll I conducted, I asked respondents whether they were familiar with each of 15 political figures in the community. As a methodological exercise, I made up a name: Tom Sakumoto. In response, 9 percent of the respondents said they were familiar with him. Of those respondents familiar with him, about half reported seeing him on television and reading about him in the newspapers.

When you obtain responses to fictitious issues, you can disregard those responses. But when the issue is real, you may have no way of telling which responses genuinely reflect attitudes and which reflect meaningless answers to an irrelevant question.

Ideally, we would like respondents to simply report that they don't know, have no opinion, or are undecided in those instances where that is the case. Unfortunately, however, they often make up answers.

Short Items Are Best

In the interests of being unambiguous and precise and of pointing to the relevance of an issue, researchers tend to create long and complicated items. That should be avoided. Respondents are often unwilling to study an item in order to understand it. The respondent should be able to read an item quickly, understand its intent, and select or provide an answer without difficulty. In general, assume that respondents will read items quickly and give quick answers. Accordingly, provide clear, short items that will not be misinterpreted under those conditions.

Avoid Negative Items

The appearance of a negation in a questionnaire item paves the way for easy misinterpretation. Asked to agree or disagree with the statement "The United States should not recognize Cuba," a sizable portion of the respondents will read over the word *not* and answer on that basis. Thus, some will agree with the statement when they're in favor of recognition, and others will agree when they oppose it. And you may never know which are which.

Similar considerations apply to other "negative" words. In a study of support for civil liberties, for example, respondents were asked whether they felt "the following kinds of people should be *prohibited* from teaching in public schools" and were presented with a list including such items as a Communist, a Ku Klux Klansman, and so forth. The response categories "yes" and "no" were given beside each entry. A comparison of the responses to this item with other items reflecting support for civil liberties strongly suggested that many respondents gave the answer "yes" to indicate willingness for such a person to teach, rather than to indicate

that such a person should be prohibited from teaching. (A later study in the series using the answer categories "permit" and "prohibit" produced much clearer results.)

In 1993 a national survey commissioned by the American Jewish Committee produced shocking results: One American in 5 believed that the Nazi Holocaust—in which 6 million Jews were reportedly killed—never happened; further, 1 in 3 Americans expressed some doubt that it had occurred. This research finding suggested that the Holocaust revisionist movement in America was powerfully influencing public opinion ("1 in 5 Polled Voices Doubt on Holocaust" 1993).

In the aftermath of this shocking news, researchers reexamined the actual question that had been asked: "Does it seem possible or does it seem impossible to you that the Nazi extermination of the Jews never happened?" On reflection, it seemed clear that the complex, double-negative question could have confused some respondents.

A new survey was commissioned and asked, "Does it seem possible to you that the Nazi extermination of the Jews never happened, or do you feel certain that it happened?" In the follow-up survey, only 1 percent of the respondents believed the Holocaust never happened, and another 8 percent said they weren't sure ("Poll on Doubt of Holocaust Is Corrected" 1994).

Avoid Biased Items and Terms

Recall from our discussion of conceptualization and operationalization in Chapter 5 that there are no ultimately true meanings for any of the concepts we typically study in social science. *Prejudice* has no ultimately correct definition; whether a given person is prejudiced depends on our definition of that term. The same general principle applies to the responses we get from people completing a questionnaire.

The meaning of someone's response to a question depends in large part on its wording. This is true of every question and answer. Some questions seem to encourage particular responses more than other questions do. In the context of questionnaires, **bias** refers to any property of questions that encourages respondents to answer in a particular way.

Most researchers recognize the likely effect of a leading question that begins, "Don't you agree

with the president of the United States that . . ." No reputable researcher would use such an item. Unfortunately, the biasing effect of items and terms is far subtler than this example suggests.

The mere identification of an attitude or position with a prestigious person or agency can bias responses. The item "Do you agree or disagree with the recent Supreme Court decision that . . ." would have a similar effect. Such wording may not produce consensus or even a majority in support of the position identified with the prestigious person or agency, but it will likely increase the level of support over what would have been obtained without such identification.

Sometimes the impact of different forms of question wording is relatively subtle. For example, when Kenneth Rasinski (1989) analyzed the results of several General Social Survey (GSS) studies of attitudes toward government spending, he found that the way programs were identified had an impact on the amount of public support they received. Here are some comparisons:

More Support	Less Support
"Assistance to the poor"	"Welfare"
"Halting rising crime rate"	"Law enforcement"
"Dealing with drug addiction"	"Drug rehabilitation"
"Solving problems of big cities"	"Assistance to big cities"
"Improving conditions of blacks"	"Assistance to blacks"
"Protecting Social Security"	"Social Security"

In 1986, for example, 62.8 percent of the respondents said too little money was being spent on "assistance to the poor," whereas in a matched survey that year, only 23.1 percent said we were spending too little on "welfare."

In this context, be wary of what researchers call the *social desirability* of questions and answers. Whenever we ask people for information, they answer through a filter of what will make them look good. This is especially true if they're interviewed face-to-face. Thus, for

> **bias** That quality of a measurement device that tends to result in a misrepresentation of what is being measured in a particular direction. For example, the questionnaire item "Don't you agree that the president is doing a good job?" would be biased in that it would generally encourage more favorable responses.

example, during the 2008 Democratic primary, many voters who might have been reluctant to vote for an African American (Barack Obama) or a woman (Hillary Clinton) might have also been reluctant to admit their racial or gender prejudice to a survey interviewer. (Some, to be sure, were not reluctant to say how they felt.)

Frauke Kreuter, Stanley Presser, and Roger Tourangeau (2008) conducted an experiment on the impact of other data-collection techniques concerning respondents' willingness to provide sensitive information that might not reflect positively on themselves—such as failing a class or being put on academic probation. Of the three methods tested, respondents were least likely to volunteer such information when interviewed in a conventional telephone interview. They were somewhat more willing when interviewed by an interactive recording, and they were most likely to provide such information when questioned in a web survey.

The best way to guard against this problem is to imagine how you would feel giving each of the answers you intend to offer to respondents. If you would feel embarrassed, perverted, inhumane, stupid, irresponsible, or otherwise socially disadvantaged by any particular response, give serious thought to how willing others will be to provide those answers.

The biasing effect of particular wording is often difficult to anticipate. For example, in both surveys and experiments, researchers sometimes ask respondents to consider hypothetical situations and say how they think they would behave. Because those constructions often involve other people, however, the names used can affect responses. For instance, researchers have long known that male names for such hypothetical people can produce different responses than female names do. Research by Joseph Kasof (1993) points to the importance of what the specific names are: whether they generally evoke positive or negative images in terms of attractiveness, age, intelligence, and so forth. Kasof's review of past research suggests there has been a tendency to use more-positively-valued names for men than for women.

The Center for Disease Control (Choi and Pak 2005) has provided an excellent analysis of various ways in which the choice of terms can bias and otherwise confuse responses to questionnaires. Among other things, they warn against using ambiguous, technical, uncommon, or vague words. Their thorough analysis provides many concrete illustrations.

As in all other research, carefully examine the purpose of your inquiry and construct items that will be most useful to it. You should never be misled into thinking there are ultimately "right" and "wrong" ways of asking the questions. Moreover, when in doubt about the best question to ask, remember that you should ask *more* than one.

These, then, are some general guidelines for writing questions to elicit data for analysis and interpretation. Next we look at how to construct questionnaires.

Questionnaire Construction

Questionnaires are used in connection with many modes of observation in social research. Although structured questionnaires are essential to and most directly associated with survey research, they are also widely used in experiments, field research, and other data-collection activities. For this reason, questionnaire construction can be an important practical skill for researchers. As we discuss the established techniques for constructing questionnaires, let's begin with some issues of questionnaire format.

General Questionnaire Format

The format of a questionnaire is just as important as the nature and wording of the questions asked. An improperly laid out questionnaire can lead respondents to miss questions, confuse them about the nature of the data desired, and even lead them to throw the questionnaire away.

As a general rule, a questionnaire should be adequately spaced and have an uncluttered layout. If a self-administered questionnaire is being designed, inexperienced researchers tend to fear that their questionnaire will look too long; as a result, they squeeze several questions onto a single line, abbreviate questions, and try to use as few pages as possible. These efforts are ill-advised and even dangerous. Putting more than one question on a line will cause some respondents to miss the second question altogether. Some respondents will misinterpret abbreviated questions. More generally, respondents who find they have spent considerable time on the first page of what seemed like a short

questionnaire will be more demoralized than respondents who quickly complete the first several pages of what initially seemed like a rather long form. Moreover, the latter will have made fewer errors and will not have been forced to reread confusing, abbreviated questions. Nor will they have been forced to write a long answer in a tiny space.

Similar problems can arise for interviewers in a face-to-face or telephone interview. Like respondents to a self-administered questionnaire, interviewers may miss questions, lose their place, and generally become frustrated and flustered. Interview questionnaires need to be formatted in a way that supports the interviewer's work, and must include any special instructions and guidelines that go beyond what respondents to a self-administered questionnaire would need.

The desirability of spreading out questions in the questionnaire cannot be overemphasized. Squeezed-together questionnaires are disastrous, whether they are to be completed by the respondents themselves or administered by trained interviewers. The processing of such questionnaires is another nightmare; I'll have more to say about that in Chapter 14.

Formats for Respondents

In one of the most common types of questionnaire items, the respondent is expected to check one response from a series. For this purpose my experience has been that boxes adequately spaced apart are the best format. Word processing makes the use of boxes a practical technique these days; setting boxes in type can be accomplished easily and neatly. You can approximate boxes by using brackets: [ ]. Even better, a few extra minutes on the computer will let you find or create genuine boxes that will give your questionnaire a more professional look. Here are some easy examples:

☐　○　❏

Rather than providing boxes to be checked, you might print a code number beside each response and ask the respondent to circle the appropriate number (see Figure 9-1). This method has the added advantage of specifying the code number to be entered later in the processing stage (see Chapter 14). If numbers are to be circled, however, you should provide clear

Did you happen to vote in the last presidential election?

  ① Yes
  2. No
  3. Don't know

Have you ever felt you were the victim of sexual discrimination?

  1. Yes
  ② No
  3. Don't know

FIGURE 9-1

Circling the Answer

© Cengage Learning®

and prominent instructions to the respondent, because many will be tempted to cross out the appropriate number, which makes data processing more difficult. (Note that the technique can be used more safely when interviewers administer the questionnaires, because the interviewers themselves record the responses.)

Contingency Questions

Quite often in questionnaires, certain questions will be relevant to some of the respondents and irrelevant to others. In a study of birth control methods, for instance, you would probably not want to ask men if they take birth control pills.

This sort of situation often arises when researchers wish to ask a series of questions about a certain topic. You may want to ask whether your respondents belong to a particular organization and, if so, how often they attend meetings, whether they have held office in the organization, and so forth. Or, you might want to ask whether respondents have heard anything about a certain political issue and then learn the attitudes of those who have heard of it.

Each subsequent question in series such as these is called a **contingency question**: Whether it is to be asked and answered is contingent on responses to the first question in the series. The

contingency question　A survey question intended for only some respondents, determined by their responses to some other question. For example, all respondents might be asked whether they belong to the Cosa Nostra, and only those who said yes would be asked how often they go to company meetings and picnics. The latter would be a contingency question.

23. Have you ever smoked marijuana?

☐ Yes
☐ No

> **If yes:** About how many times have you smoked marijuana?
>
> ☐ Once
> ☐ 2 to 5 times
> ☐ 6 to 10 times
> ☐ 11 to 20 times
> ☐ More than 20 times

FIGURE 9-2

Contingency Question Format. Contingency questions offer a structure for exploring subject areas logically in some depth.

© Cengage Learning®

24. Have you ever been abducted by aliens?

☐ Yes
☐ No

> **If yes:** Did they let you steer the ship?
>
> ☐ Yes
> ☐ No
>
> > **If yes:** How fast did you go?
> >
> > ☐ Warp speed
> > ☐ Weenie speed

FIGURE 9-3

Contingency Table. Sometimes it will be appropriate for certain kinds of respondents to skip over inapplicable questions. To avoid confusion, you should be sure to provide clear instructions to that end.

© Cengage Learning®

proper use of contingency questions can facilitate the respondents' task in completing the questionnaire, because they are not faced with trying to answer questions irrelevant to them.

There are several formats for contingency questions. The one shown in Figure 9-2 is probably the clearest and most effective. Note two key elements in this format. First, the contingency question is isolated from the other questions by being set off to the side and enclosed in a box. Second, an arrow connects the contingency question to the answer on which it is contingent. In the illustration, only those respondents answering yes are expected to answer the contingency question. The rest of the respondents should simply skip it.

Note that the questions shown in Figure 9-2 could have been dealt with in a single question. The question might have read, "How many times, if any, have you smoked marijuana?" The response categories, then, might have read: "Never," "Once," "2 to 5 times," and so forth. This single question would apply to all respondents, and each would find an appropriate answer category. Such a question, however, might put some pressure on respondents to report having smoked marijuana, because the main question asks how many times they have smoked it, even though it allows for those *exceptional* cases *who have never smoked marijuana even once.* (The emphases used in the previous sentence give a fair indication of how respondents might read the question.) The contingency question format illustrated in Figure 9-2 should reduce the subtle pressure on respondents to report having smoked marijuana.

Used properly, even rather complex sets of contingency questions can be constructed without confusing the respondent. Figure 9-3 illustrates a more complicated example.

Sometimes a set of contingency questions is long enough to extend over several pages. Suppose you're studying political activities of college students, and you wish to ask a large number of questions of those students who have voted in a national, state, or local election. You could separate out the relevant respondents with an initial question such as "Have you ever voted in a national, state, or local election?" but it would be confusing to place the contingency questions in a box stretching over several pages. It would make more sense to enter instructions, in parentheses after each answer, telling respondents to answer or skip the contingency questions. Figure 9-4 provides an illustration of this method.

In addition to these instructions, it's worthwhile to place additional directions at the top of each page containing only the contingency questions. For example, you might say, "This page is only for respondents who have voted in a

13. Have you ever voted in a national, state, or local election?

☐ Yes (Please answer questions 14–25.)
☐ No (Please skip questions 14–25.
     Go directly to question 26 on page 8.)

FIGURE 9-4

Instructions to Skip

© Cengage Learning®

national, state, or local election." Clear guidelines such as these spare respondents the frustration of reading and puzzling over questions irrelevant to them and increase the likelihood of responses from those for whom the questions are relevant.

Matrix Questions

Quite often, you'll want to ask several questions that have the same set of answer categories. This is typically the case whenever the Likert response categories are used. In such cases, it is often possible to construct a matrix of items and answers as illustrated in Figure 9-5.

This format offers several advantages over other formats. First, it uses space efficiently. Second, respondents will probably find it faster to complete a set of questions presented in this fashion than in other ways. In addition, this format may increase the comparability of responses given to different questions for the respondent as well as for the researcher. Because respondents can quickly review their answers to earlier items in the set, they might choose between, say, "strongly agree" and "agree" on a given statement by comparing the strength of their agreement with their earlier responses in the set.

There are some dangers inherent in using this format, however. Its advantages may encourage you to structure an item so that the responses fit into the matrix format when a different, more idiosyncratic set of responses might be more appropriate. Also, the matrix question format can foster a response-set among some respondents: They may develop a pattern of, say, agreeing with all the statements. This would be especially likely if the set of statements began with several that indicated a particular orientation (for example, a liberal political perspective) with only a few later ones representing the opposite orientation. Respondents might assume that all the statements represented the same orientation and, reading quickly, misread some of them, thereby giving the wrong answers. This problem can be reduced somewhat by alternating statements representing different orientations and by making all statements short and clear.

Ordering Items in a Questionnaire

The order in which questionnaire items are presented can also affect responses. First, the appearance of one question can affect the answers given to later ones. For example, if several questions have been asked about the dangers of terrorism to the United States and then a question asks respondents to volunteer (open-endedly) what they believe to represent dangers to the United States, terrorism will receive more citations than would otherwise be the case. In this situation, it's preferable to ask the open-ended question first.

Similarly, if respondents are asked to assess their overall religiosity ("How important is your religion to you in general?"), their responses to later questions concerning specific aspects of religiosity will be aimed at consistency with the prior assessment. The converse is true as well. If respondents are first asked specific questions about different aspects of their religiosity, their subsequent overall assessment will reflect the earlier answers. The order of responses within a question can also make a difference (Bishop and Smith 2001).

17. Beside each of the statements presented below, please indicate whether you Strongly Agree (**SA**), Agree (**A**), Disagree (**D**), Strongly Disagree (**SD**), or are Undecided (**U**).

	SA	A	D	SD	U
a. What this country needs is more law and order..........	☐	☐	☐	☐	☐
b. The police should be disarmed in America...........	☐	☐	☐	☐	☐
c. During riots, looters should be shot on sight.........	☐	☐	☐	☐	☐
etc.					

FIGURE 9-5

Matrix Question Format. Matrix questions offer an efficient format for presenting a set of closed-ended questionnaire items that have the same response categories.

The impact of item order is not uniform. When J. Edwin Benton and John Daly (1991) conducted a local government survey, they found that the less-educated respondents were more influenced by the order of questionnaire items than those with more education were.

Some researchers attempt to overcome this effect by randomizing the order of items. This effort is usually futile. In the first place, a randomized set of items will probably strike respondents as chaotic and worthless. The random order also makes it more difficult for respondents to answer, because they must continually switch their attention from one topic to another. Finally, even a randomized ordering of items will have the effect discussed previously—except that you'll have no control over the effect.

The safest solution is sensitivity to the problem. Although you cannot avoid the effect of item order, try to estimate what that effect will be so that you can interpret results meaningfully. If the order of items seems especially important in a given study, you might construct more than one version of the questionnaire with different orderings of the items. You will then be able to determine the effects by comparing responses to the various versions. At the very least, you should pretest your questionnaire in the different forms. (We'll discuss pretesting in a moment.)

The desired ordering of items differs between interviews and self-administered questionnaires. In the latter, it's usually best to begin the questionnaire with the most interesting set of items. The potential respondents who glance casually over the first few items should want to answer them. Perhaps the items will ask for attitudes they're aching to express. At the same time, however, the initial items should not be threatening. (It might be a bad idea to begin with items about sexual behavior or drug use.) Requests for duller, demographic data (age, sex, and the like) should generally be placed at the end of a self-administered questionnaire. Placing these items at the beginning, as many inexperienced researchers are tempted to do, gives the questionnaire the initial appearance of a routine form, and the person receiving it may not be motivated to complete it.

Just the opposite is generally true for interview surveys. When the potential respondent's door first opens, the interviewer must gain rapport quickly. After a short introduction to the study, the interviewer can best begin by enumerating the members of the household, getting demographic data about each. Such items are easily answered and generally nonthreatening. Once the initial rapport has been established, the interviewer can then move into the area of attitudes and more-sensitive matters. An interview that began with the question "Do you believe in witchcraft?" would probably end rather quickly (though hopefully not in a puff of smoke).

Questionnaire Instructions

Every questionnaire, whether it is to be completed by respondents or administered by interviewers, should contain clear instructions and introductory comments where appropriate.

It's useful to begin every self-administered questionnaire with basic instructions for completing it. Although many people these days have experience with forms and questionnaires, begin by telling them exactly what you want: that they are to indicate their answers to certain questions by placing a check mark or an **X** in the box beside the appropriate answer or by writing in their answer when asked to do so. If many open-ended questions are used, respondents should be given some guidelines about whether brief or lengthy answers are expected. If you wish to encourage your respondents to elaborate on their responses to closed-ended questions, that should be noted.

If a questionnaire has subsections—political attitudes, religious attitudes, background data—introduce each with a short statement concerning its content and purpose. For example, "In this section, we would like to know what people consider to be the most important community problems." Demographic items at the end of a self-administered questionnaire might be introduced thus: "Finally, we would like to know just a little about you so we can see how different types of people feel about the issues we have been examining."

Short introductions and explanations such as these help the respondent make sense of the questionnaire. They make the questionnaire seem less chaotic, especially when it taps a variety of data. And they help put the respondent in the proper frame of mind for answering the questions.

Some questions may require special instructions to facilitate proper answering. This is especially true if a given question varies from the

general instructions pertaining to the whole questionnaire. Some specific examples will illustrate this situation.

Despite attempts to provide mutually exclusive answers in closed-ended questions, often more than one answer will apply for respondents. If you want a single answer, you should make this perfectly clear in the question. An example would be "From the list below, please check the primary reason for your decision to attend college." Often the main question can be followed by a parenthetical note: "Please check the one best answer." If, on the other hand, you want the respondent to check as many answers as apply, you should make this clear.

When the respondent is supposed to rank-order a set of answer categories, the instructions should indicate this, and a different type of answer format should be used (for example, blanks instead of boxes). These instructions should indicate how many answers are to be ranked (for example: all; only the first and second; only the first and last; the most important and least important). These instructions should also spell out the order of ranking (for example: "Place a 1 beside the most important item, a 2 beside the next most important, and so forth"). Rank-ordering of responses is often difficult for respondents, however, because they may have to read and reread the list several times, so this technique should be used only in those situations where no other method will produce useful data.

In multiple-part matrix questions, giving special instructions is useful unless the same format is used throughout the questionnaire. Sometimes respondents will be expected to check one answer in each column of the matrix; in other questionnaires they'll be expected to check one answer in each row. Whenever the questionnaire contains both formats, it's useful to add an instruction clarifying which is expected in each case.

Pretesting the Questionnaire

No matter how carefully researchers design a data-collection instrument such as a questionnaire, there is always the possibility—indeed the certainty—of error. They will always make some mistake: write an ambiguous question, or one that people cannot answer, or commit some other violation of the rules just discussed.

The surest protection against such errors is to pretest the questionnaire in full or in part. Give the questionnaire to the 10 people in your bowling league, for example. It's not usually essential that the pretest subjects comprise a representative sample, although you should use people for whom the questionnaire is at least relevant.

By and large, it's better to ask people to complete the questionnaire than to read through it looking for errors. All too often, a question seems to make sense on a first reading, but it proves to be impossible to answer.

Stanley Presser and Johnny Blair (1994) describe several different pretesting strategies and report on the effectiveness of each. They also provide data on the cost of the various methods. Paul Beatty and Gordon Willis (2007) offer a useful review of "cognitive interviewing." In this technique, the pretest includes gathering respondents' comments about the questionnaire itself, so that the researchers can see which questions are communicating effectively and collecting the information sought.

There are many more tips and guidelines for questionnaire construction, but covering them all would take a book in itself. For now, I'll complete this discussion with an illustration of a real questionnaire, showing how some of these comments find substance in practice.

Before turning to the illustration, however, I want to mention a critical aspect of questionnaire design: precoding. Because the information collected by questionnaires is typically transformed into some type of computer format, it's usually appropriate to include data-processing instructions on the questionnaire itself. These instructions indicate where specific pieces of information will be stored in the machine-readable data files. Notice that the following illustration has been precoded with the mysterious numbers that appear near questions and answer categories.

A Composite Illustration

Figure 9-6 is part of a questionnaire used by the University of Chicago's National Opinion Research Center in its General Social Survey. The questionnaire dealt with people's attitudes toward the government and was designed to be self-administered, though most of the GSS is conducted in face-to-face interviews.

10. Here are some things the government might do for the economy. Circle one number for each action to show whether you are in favor of it or against it.

> 1. **Strongly in favor of**
> 2. **In favor of**
> 3. **Neither in favor of nor against**
> 4. **Against**
> 5. **Strongly against**

PLEASE CIRCLE A NUMBER

a. Control of wages by legislation	1	2	3	4	5	28/
b. Control of prices by legislation	1	2	3	4	5	29/
c. Cuts in government spending	1	2	3	4	5	30/
d. Government financing of projects to create new jobs	1	2	3	4	5	31/
e. Less government regulation of business	1	2	3	4	5	32/
f. Support for industry to develop new products and technology	1	2	3	4	5	33/
g. Supporting declining industries to protect jobs	1	2	3	4	5	34/
h. Reducing the work week to create more jobs	1	2	3	4	5	35/

11. Listed below are various areas of government spending. Please indicate whether you would like to see more or less government spending in each area. Remember that if you say "much more," it might require a tax increase to pay for it.

> 1. **Spend much more**
> 2. **Spend more**
> 3. **Spend the same as now**
> 4. **Spend less**
> 5. **Spend much less**
> 8. **Can't choose**

PLEASE CIRCLE A NUMBER

a. The environment	1	2	3	4	5	8	36/
b. Health	1	2	3	4	5	8	37/
c. The police and law enforcement	1	2	3	4	5	8	38/
d. Education	1	2	3	4	5	8	39/
e. The military and defense	1	2	3	4	5	8	40/
f. Retirement benefits	1	2	3	4	5	8	41/
g. Unemployment benefits	1	2	3	4	5	8	42/
h. Culture and the arts	1	2	3	4	5	8	43/

12. If the government *had* to choose between keeping down inflation or keeping down unemployment, to which do you think it should give highest priority?

Keeping down inflation	1	44/
Keeping down unemployment	2	
Can't choose	8	

13. Do you think that labor unions in this country have too much power or too little power?

Far too much power	1	45/
Too much power	2	
About the right amount of power	3	
Too little power	4	
Far too little power	5	
Can't choose	8	

FIGURE 9-6

A Sample Questionnaire. This questionnaire excerpt is from the General Social Survey, a major source of data for analysis by social researchers around the world.
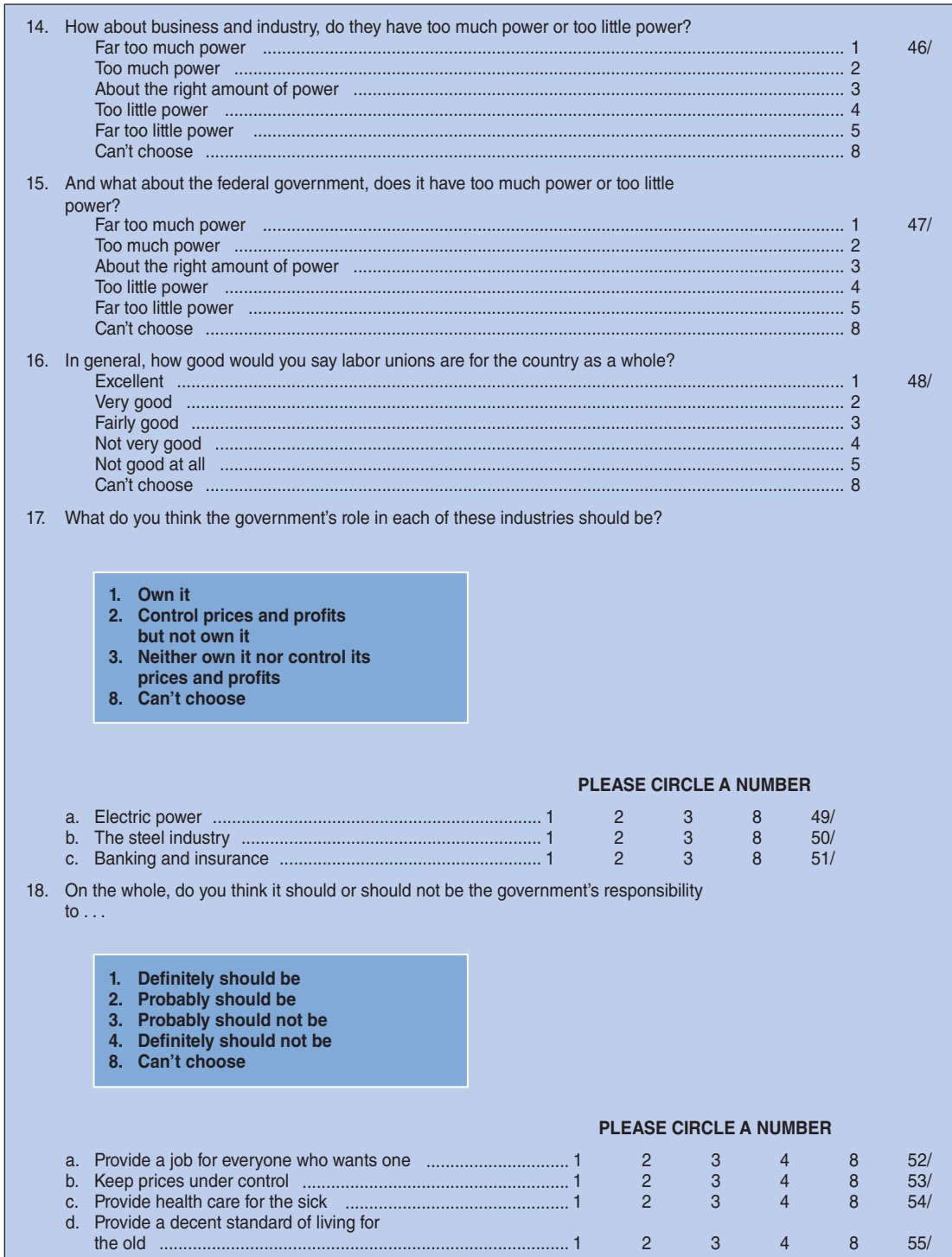
14. How about business and industry, do they have too much power or too little power?
    Far too much power 1   46/
    Too much power 2
    About the right amount of power 3
    Too little power 4
    Far too little power 5
    Can't choose 8

15. And what about the federal government, does it have too much power or too little power?
    Far too much power 1   47/
    Too much power 2
    About the right amount of power 3
    Too little power 4
    Far too little power 5
    Can't choose 8

16. In general, how good would you say labor unions are for the country as a whole?
    Excellent 1   48/
    Very good 2
    Fairly good 3
    Not very good 4
    Not good at all 5
    Can't choose 8

17. What do you think the government's role in each of these industries should be?

> 1. **Own it**
> 2. **Control prices and profits but not own it**
> 3. **Neither own it nor control its prices and profits**
> 8. **Can't choose**

PLEASE CIRCLE A NUMBER

a. Electric power 1   2   3   8   49/
b. The steel industry 1   2   3   8   50/
c. Banking and insurance 1   2   3   8   51/

18. On the whole, do you think it should or should not be the government's responsibility to . . .

> 1. **Definitely should be**
> 2. **Probably should be**
> 3. **Probably should not be**
> 4. **Definitely should not be**
> 8. **Can't choose**

PLEASE CIRCLE A NUMBER

a. Provide a job for everyone who wants one 1   2   3   4   8   52/
b. Keep prices under control 1   2   3   4   8   53/
c. Provide health care for the sick 1   2   3   4   8   54/
d. Provide a decent standard of living for the old 1   2   3   4   8   55/

FIGURE 9-6

(*Continued*)

Self-Administered Questionnaires

So far we've discussed how to formulate questions and how to design effective questionnaires. As important as these tasks are, the labor will be wasted unless the questionnaire produces useful data—which means that respondents actually complete the questionnaire. We turn now to the major methods for getting responses to questionnaires.

I've referred several times in this chapter to interviews and self-administered questionnaires. Actually, there are three main methods of administering survey questionnaires to a sample of respondents: self-administered questionnaires, in which respondents are asked to complete the questionnaire themselves; surveys administered by interviewers in face-to-face encounters; and surveys conducted by telephone. This section and the next two discuss each of these methods in turn. A fourth section addresses online surveys, a new technique rapidly growing in popularity.

The most common form of self-administered questionnaire is the mail survey. However, there are several other techniques that are often used as well. At times, it may be appropriate to administer a questionnaire to a group of respondents gathered at the same place at the same time. For example, a survey of students taking introductory psychology might be conducted during class. High school students might be surveyed during homeroom period.

Some recent experimentation has been conducted with regard to the home delivery of questionnaires. A research worker delivers the questionnaire to the home of sample respondents and explains the study. Then the questionnaire is left for the respondent to complete, and the researcher picks it up later.

Home delivery and the mail can also be used in combination. Questionnaires are mailed to families, and then research workers visit homes to pick up the questionnaires and check them for completeness. Just the opposite technique is to have questionnaires hand-delivered by research workers with a request that the respondents mail the completed questionnaires to the research office.

On the whole, when a research worker either delivers the questionnaire, picks it up, or both, the completion rate seems higher than it

is for straightforward mail surveys. Additional experimentation with this technique is likely to point to other ways to improve completion rates while reducing costs. The remainder of this section, however, is devoted specifically to the mail survey, which is still the typical form of self-administered questionnaire.

Mail Distribution and Return

The basic method for collecting data through the mail has been to send a questionnaire accompanied by a letter of explanation and a self-addressed, stamped envelope for returning the questionnaire. The respondent is expected to complete the questionnaire, put it in the envelope, and return it. If, by any chance, you've received such a questionnaire and failed to return it, it would be valuable to recall the reasons you had for not returning it and keep them in mind any time you plan to send questionnaires to others.

A common reason for not returning questionnaires is that it's too much trouble. To overcome this problem, researchers have developed several ways to make returning them easier. For instance, a self-mailing questionnaire requires no return envelope: When the questionnaire is folded a particular way, the return address appears on the outside. The respondent therefore doesn't have to worry about losing the envelope.

More-elaborate designs are available also. The university student questionnaire to be described later in this chapter was bound in a booklet with a special, two-panel back cover. Once the questionnaire was completed, the respondent needed only to fold out the extra panel, wrap it around the booklet, and seal the whole thing with the adhesive strip running along the edge of the panel. The foldout panel contained my return address and postage. When I repeated the study a couple of years later, I improved on the design. Both the front and back covers had foldout panels: one for sending the questionnaire out and the other for getting it back—thus avoiding the use of envelopes altogether.

The point here is that anything you can do to make the job of completing and returning the questionnaire easier will improve your study. Imagine receiving a questionnaire that made no provisions for its return to the researcher.

Suppose you had to (1) find an envelope, (2) write the address on it, (3) figure out how much postage it required, and (4) put the stamps on it. How likely is it that you would return the questionnaire?

A few brief comments on postal options are in order. You have options for mailing questionnaires out and for getting them returned. On outgoing mail, your choices are essentially between first-class postage and bulk rate. First class is more certain, but bulk rate is far cheaper. (Check your local post office for rates and procedures.) On return mail, your choice is between postage stamps and business-reply permits. Here, the cost differential is more complicated. If you use stamps, you pay for them whether people return their questionnaires or not. With the business-reply permit, you pay for only those that are used, but you pay an additional surcharge of about a nickel. This means that stamps are cheaper if a lot of questionnaires are returned, but business-reply permits are cheaper if fewer are returned (and there is no way for you to know in advance how many will be returned).

There are many other considerations involved in choosing among the several postal options. Some researchers, for example, feel that using postage stamps communicates more "humanness" and sincerity than using bulk rate and business-reply permits does. Others worry that respondents will peel off the stamps and use them for some purpose other than returning the questionnaires. Because both bulk rate and business-reply permits require establishing accounts at the post office, you'll probably find stamps much easier for small surveys.

Monitoring Returns

The mailing of questionnaires sets up a new research question that may prove valuable to a study. Researchers shouldn't sit back idly as questionnaires are returned; instead, they should undertake a careful recording of the varying rates of return among respondents.

An invaluable tool in this activity is a return-rate graph. The day on which questionnaires were mailed is labeled Day 1 on the graph, and on every day thereafter the number of returned questionnaires is logged on the graph. It's usually best to compile two graphs. One shows the

number returned each day—rising over time, then dropping. The second reports the cumulative number or percentage. In part, this activity provides the researchers with gratification, as they get to draw a picture of their successful data collection. More important, however, it serves as their guide to how the data collection is going. If follow-up mailings are planned, the graph provides a clue about when such mailings should be launched. (The dates of subsequent mailings also should be noted on the graph.)

As completed questionnaires are returned, each should be opened, scanned, and assigned an identification (ID) number. These numbers should be assigned serially as the questionnaires are returned, even if other identification numbers have already been assigned. Two examples should illustrate the important advantages of this procedure.

Let's assume you're studying attitudes toward a political figure. In the middle of the data collection, the media break the story that the politician is having extramarital affairs. By knowing the date of that public disclosure and the dates when questionnaires were received, you'll be in a position to determine the effects of the disclosure. (Recall from Chapter 8 the discussion of history in connection with experiments.)

In a less sensational way, serialized ID numbers can be valuable in estimating non-response biases in the survey. Barring more-direct tests of bias, you may wish to assume that those who failed to answer the questionnaire will be more like respondents who delayed answering than like those who answered right away. An analysis of questionnaires received at different points in the data collection might then be used for estimates of sampling bias. For example, if the grade point averages (GPAs) reported by student respondents decrease steadily through the data collection, with those replying right away having higher GPAs and those replying later having lower GPAs, you might tentatively conclude that those who failed to answer at all have lower GPAs yet. Although it would not be advisable to make statistical estimates of bias in this fashion, you could take advantage of approximate estimates based on the patterns you've observed.

If respondents have been identified for purposes of follow-up mailing, then preparations for those mailings should be made as the

questionnaires are returned. The case study later in this section discusses this process in greater detail.

Follow-Up Mailings

Follow-up mailings may be administered in several ways. In the simplest, non-respondents are simply sent a letter of additional encouragement to participate. A better method, however, is to send a new copy of the survey questionnaire with the follow-up letter. If potential respondents have not returned their questionnaires after two or three weeks, the questionnaires have probably been lost or misplaced. Receiving a follow-up letter might encourage them to look for the original questionnaire, but if they can't find it easily, the letter may go for naught.

The methodological literature strongly suggests that follow-up mailings provide an effective method for increasing return rates in mail surveys. In general, the longer a potential respondent delays replying, the less likely he or she is to do so at all. Properly timed follow-up mailings, then, provide additional stimuli to respond.

The effects of follow-up mailings will be seen in the response-rate curves recorded during data collection. The initial mailings will be followed by a rise and subsequent subsiding of returns; the follow-up mailings will spur a resurgence of returns; and more follow-ups will do the same. In practice, three mailings (an original and two follow-ups) seem the most efficient.

The timing of follow-up mailings is also important. Here the methodological literature offers less-precise guides, but I've found that two or three weeks is a reasonable space between mailings. (This period might be increased by a few days if the mailing time—out and in—is more than two or three days.)

If the individuals in the survey sample are not identified on the questionnaires, it may not be possible to remail only to non-respondents. In such a case, send your follow-up mailing to all members of the sample, thanking those who may have already participated and encouraging those who have not to do so. (The case study reported later describes yet another method you can use in an anonymous mail survey.)

Response Rates

A question that new survey researchers frequently ask concerns the percentage return rate, or the response rate, that should be achieved in a survey. The body of inferential statistics used in connection with survey analysis assumes that *all* members of the initial sample complete the survey. Because this almost never happens, non-response bias becomes a concern, with the researcher testing (and hoping) for the possibility that the respondents look essentially like a random sample of the initial sample, and thus a somewhat smaller random sample of the total population.

Nevertheless, overall **response rate** is one guide to the representativeness of the sample respondents. If a high response rate is achieved, there is less chance of significant non-response bias than with a low rate. Conversely, a low response rate is a danger signal, because the non-respondents are likely to differ from the respondents in ways other than just their willingness to participate in the survey. Richard Bolstein (1991), for example, found that those who did not respond to a pre-election political poll were less likely to vote than were those who did participate. Estimating the turnout rate from just the survey respondents, then, would have overestimated the number who would show up at the polls. Ironically, of course, since the non-respondents were unlikely to vote, the preferences of the survey participants might offer a good estimate of the election results.

In the book *Standard Definitions,* the American Association for Public Opinion Research (AAPOR 2008: 4–5) defines the response rate, and further distinguishes contact rates, refusal rates, and cooperation rates.

- Response rates—The number of complete interviews with reporting units divided by the number of eligible reporting units in the sample. The report provides six definitions of response rates, ranging from the definition that yields the lowest

response rate The number of people participating in a survey divided by the number selected in the sample, in the form of a percentage. This is also called the *completion rate* or, in self-administered surveys, the *return rate*: the percentage of questionnaires sent out that are returned.

rate to the definition that yields the highest rate, depending on how partial interviews are considered and how cases of unknown eligibility are handled.

- Cooperation rates—The proportion of all cases interviewed of all eligible units ever contacted. The report provides four definitions of cooperation rates, ranging from a minimum or lowest rate, to a maximum or highest rate.
- Refusal rates—The proportion of all cases in which a housing unit or the respondent refuses to be interviewed, or breaks-off an interview, of all potentially eligible cases. The report provides three definitions of refusal rates, which differ in the way they treat dispositions of cases of unknown eligibility.
- Contact rates—The proportion of all cases in which some responsible housing unit member was reached. The report provides three definitions of contact rates.

While response rates logically affect the quality of survey data, this is not always in fact the case, as Robert Groves (2006) points out. With recent declines in response rates, this is a topic under careful study by survey researchers. At the same time, higher responses are a goal.

As you can imagine, one of the more persistent discussions among survey researchers concerns ways of increasing response rates. You'll recall that this was a chief concern in the earlier discussion of options for mailing out and receiving questionnaires. Survey researchers have developed many ingenious techniques addressing this problem. Some have experimented with novel formats. Others have tried paying respondents to participate. The problem with paying, of course, is that it's expensive to make meaningfully high payment to hundreds or thousands of respondents, but some imaginative alternatives have been used. Some researchers have said, "We want to get your two-cents' worth on some issues, and we're willing to pay"—enclosing two pennies. Another enclosed a quarter, suggesting that the respondent make some little child happy. Still others have enclosed paper money. Similarly, Michael Davern and his colleagues (2003) found that financial incentives also increased completion rates in face-to-face interview surveys (discussed in the next section).

Don Dillman (2007) has spent decades painstakingly assessing the various techniques that survey researchers have used to increase return rates on mail surveys, and he evaluates the impact of each. More important, Dillman stresses the necessity of paying attention to all aspects of the study—what he calls the "Tailored Design Method"—rather than one or two special gimmicks.

Having said all this, there is no absolutely acceptable level of response to a mail survey, except for 100 percent. While it is possible to achieve response rates of 70 percent or more, most mail surveys probably fall below that level. Thus, it's important to test for non-response bias wherever possible.

Compensation for Respondents

It is fairly common practice to pay experimental and focus group subjects for their participation, though it has been rare in other research methods. Whether to pay survey respondents is sometimes discussed and often controversial.

In addition to cash payments, researchers have sometimes employed gift certificates, contributions to charities, lotteries, and other prize drawings. In a survey of New Zealanders, Mike Brennan and Jan Charbonneau (2009) sent chocolates as an incentive for participation.

Some researchers have provided incentives to all those selected in the sample during the first contact. In the case of cash incentives in mail surveys, this means respondents get the incentive whether they participate or not. In other cases, the researchers have provided or offered incentives in follow-up contacts with non-respondents, though this creates a problem of inequity, with the most cooperative people getting no compensation.

In a 1999 review of studies of this topic, Singer, Groves, and Corning found that with very few exceptions, response rates are increased by the use of incentives in mail surveys, face-to-face interviews, and telephone polls. Also, the authors found no evidence of negative effects on the quality of responses collected. A decade later, Petrolia and Bhattacharee (2009) reviewed past experience with incentives and conducted their own study. They confirmed that incentives increase response rates, and they found that prepaid incentives had a greater effect than those introduced later in the process.

J. Michael Brick and his colleagues (2012) reported high response rates with a two-stage mail survey. This method began with an address-based sampling (ABS) of households that then received a short demographic questionnaire designed to gather relevant characteristics about their members. Next, a subsample was selected from among those identified as appropriate to the particular survey focus, and a follow-up questionnaire was then sent. Both mailings were accompanied by a $1 cash incentive, and additional phone calls and postcard reminders were used with non-respondents.

A Case Study

The steps involved in the administration of a mail survey are many and can best be appreciated in a walk-through of an actual study. Accordingly, this section concludes with a detailed description of how the student survey we discussed in Chapter 7, as an illustration of systematic sampling, was administered. This study did not represent the theoretical ideal for such studies, but in that regard it serves our present purposes all the better. The study was conducted by the students in my graduate seminar in survey research methods.

As you may recall, 1,100 students were selected from the university registration records through a stratified, systematic sampling procedure. For each student selected, six self-adhesive mailing labels were printed.

By the time we were ready to distribute the questionnaires, it became apparent that our meager research funds wouldn't cover several mailings to the entire sample of 1,100 students (questionnaire printing costs were higher than anticipated). As a result, we chose a systematic two-thirds sample of the mailing labels, yielding a subsample of 733 students.

Earlier, we had decided to keep the survey anonymous in the hope of encouraging more-candid responses to some sensitive questions. (Later surveys of the same issues among the same population indicated this anonymity was unnecessary.) Thus, the questionnaires would carry no identification of students on them. At the same time, we hoped to reduce the follow-up mailing costs by mailing only to non-respondents.

To achieve both of these aims, a special postcard method was devised. Each student was mailed a questionnaire that carried no identifying marks, plus a postcard addressed to the research office—with one of the student's mailing labels affixed to the reverse side of the card. The introductory letter asked the student to complete and return the questionnaire—assuring anonymity—and to return the postcard simultaneously. Receiving the postcard would tell us—without indicating which questionnaire it was—that the student had returned his or her questionnaire. This procedure would then facilitate follow-up mailings.

The 32-page questionnaire was printed in booklet form. The three-panel cover described earlier in this chapter permitted the questionnaire to be returned without an additional envelope.

A letter introducing the study and its purposes was printed on the front cover of the booklet. It explained why the study was being conducted (to learn how students feel about a variety of issues), how students had been selected for the study, the importance of each student's responding, and the mechanics of returning the questionnaire.

Students were assured that their responses to the survey were anonymous, and the postcard method was explained. A statement followed about the auspices under which the study was being conducted, and a telephone number was provided for those who might want more information about the study. (Five students called for information.)

By printing the introductory letter on the questionnaire, we avoided the necessity of enclosing a separate letter in the outgoing envelope, thereby simplifying the task of assembling mailing pieces.

The materials for the initial mailing were assembled as follows. (1) One mailing label for each student was stuck on a postcard. (2) Another label was stuck on an outgoing manila envelope. (3) One postcard and one questionnaire were placed in each envelope—with a glance to ensure that the name on the postcard and on the envelope were the same in each case.

The distribution of the survey questionnaires had been set up for a bulk-rate mailing. Once the questionnaires had been stuffed into envelopes, they were grouped by zip code, tied in bundles, and delivered to the post office.

Shortly after the initial mailing, questionnaires and postcards began arriving at the research office. Questionnaires were opened, scanned, and assigned identification numbers as described earlier in this chapter. For every postcard received, a search was made for that student's remaining labels, and they were destroyed.

After two or three weeks, the remaining mailing labels were used to organize a follow-up mailing. This time a special, separate letter of appeal was included in the mailing piece. The new letter indicated that many students had returned their questionnaires already, and it was very important for all others to do so as well.

The follow-up mailing stimulated a resurgence of returns, as expected, and the same logging procedures continued. The returned postcards told us which additional mailing labels to destroy. Unfortunately, time and financial pressures made a third mailing impossible, despite initial plans to do so, but the two mailings resulted in an overall return rate of 62 percent.

This illustration should give you a fairly good sense of what's involved in the execution of mailed self-administered questionnaires. Let's turn now to the second principal method of conducting surveys, in-person interviews.

Interview Surveys

The **interview** is an alternative method of collecting survey data. Rather than asking respondents to read questionnaires and enter their own answers, researchers send interviewers to ask the questions orally and to record respondents' answers. Interviewing is typically done in a face-to-face encounter, but telephone interviewing, discussed in the next section, follows most of the same guidelines.

Most interview surveys require more than one interviewer, although you might undertake a small-scale interview survey yourself. Portions of this section will discuss methods for training and supervising a staff of interviewers assisting you with a survey. Here we deal specifically with survey interviewing; Chapter 10 discusses the less-structured, in-depth interviews often conducted in qualitative field research.

The Role of the Survey Interviewer

There are several advantages to having a questionnaire administered by an interviewer rather than a respondent. To begin with, interview surveys typically attain higher response rates than mail surveys do. A properly designed and executed interview survey ought to achieve a completion rate of at least 80 to 85 percent. (Federally funded surveys often require one of these response rates.) Respondents seem more reluctant to turn down an interviewer standing on their doorstep than to throw away a mail questionnaire.

The presence of an interviewer also generally decreases the number of "don't knows" and "no answers." If minimizing such responses is important to the study, the interviewer can be instructed to probe for answers ("If you had to pick one of the answers, which do you think would come closest to your feelings?").

Further, if a respondent clearly misunderstands the intent of a question or indicates that he or she does not understand, the interviewer can clarify matters, thereby obtaining relevant responses. (As we'll discuss shortly, such clarifications must be strictly controlled through formal specifications.)

Finally, the interviewer can observe respondents as well as ask questions. For example, the interviewer can note the quality of the dwelling, the presence of various possessions, the respondent's ability to speak English, the respondent's general reactions to the study, and so forth. In one survey of students, respondents were given a short, self-administered questionnaire to complete—concerning sexual attitudes and behavior—during the course of the interview. While respondents completed the questionnaire, the interviewer made detailed notes regarding their dress and grooming.

This procedure raises an ethical issue. Some researchers have objected that such practices violate the spirit of the agreement by which the respondent has allowed the interview. Although ethical issues seldom are clear-cut in social

interview A data-collection encounter in which one person (an interviewer) asks questions of another (a respondent). Interviews may be conducted face-to-face or by telephone.

research, it's important to be sensitive to them, as we saw in Chapter 3.

Survey research is of necessity based on an unrealistic stimulus-response theory of cognition and behavior. Researchers must assume that a questionnaire item will mean the same thing to every respondent, and every given response must mean the same when given by different respondents. Although this is an impossible goal, survey questions are drafted to approximate the ideal as closely as possible.

The interviewer must also fit into this ideal situation. The interviewer's presence should affect neither a respondent's perception of a question nor the answer given. In other words, the interviewer should be a neutral medium through which questions and answers are transmitted.

As such, different interviewers should obtain exactly the same responses from a given respondent. (Recall our earlier discussions of *reliability*.) This neutrality has a special importance in area samples. To save time and money, a given interviewer is typically assigned to complete all the interviews in a particular geographic area—a city block or a group of nearby blocks. If the interviewer does anything to affect the responses obtained, the bias thus interjected might be interpreted as a characteristic of that area.

Let's suppose that a survey is being done to determine attitudes toward low-cost housing in order to help in the selection of a site for a new government-sponsored development. An interviewer assigned to a given neighborhood might—through word or gesture—communicate his or her own distaste for low-cost housing developments. Respondents might therefore tend to give responses in general agreement with the interviewer's own position. The results of the survey would indicate that the neighborhood in question strongly resists construction of the development in its area when in fact their apparent resistance simply reflects the interviewer's attitudes.

General Guidelines for Survey Interviewing

The manner in which interviews ought to be conducted will vary somewhat by survey population and survey content. Nevertheless, some general guidelines apply to most interviewing situations.

Appearance and Demeanor

As a rule, interviewers should dress in a fashion similar to that of the people they'll be interviewing. A richly dressed interviewer will probably have difficulty getting good cooperation and responses from poorer respondents; a poorly dressed interviewer will have similar difficulties with richer respondents. To the extent that the interviewer's dress and grooming differ from those of the respondents, it should be in the direction of cleanliness and neatness in modest apparel. If cleanliness is not next to godliness, it appears at least to be next to neutrality. Although middle-class neatness and cleanliness may not be accepted by all sectors of U.S. society, they remain the primary norm and are the most likely to be acceptable to the largest number of respondents.

Dress and grooming are typically regarded as signs of a person's attitudes and orientations. Torn jeans, green hair, tattoos, and razor blade earrings may communicate—correctly or incorrectly—that the interviewer is politically radical, sexually permissive, favorable to drug use, and so forth. Any of these impressions could bias responses or affect the willingness of people to be interviewed.

In demeanor, interviewers should be pleasant if nothing else. Because they'll be prying into a respondent's personal life and attitudes, they must communicate a genuine interest in getting to know the respondent, without appearing to spy. They must be relaxed and friendly, without being too casual or clinging. Good interviewers also have the ability to determine very quickly the kind of person the respondent will feel most comfortable with, the kind of person the respondent would most enjoy talking to. Clearly, the interview will be more successful in this case. Further, because respondents are asked to volunteer a portion of their time and to divulge personal information, they deserve the most enjoyable experience the researcher and interviewer can provide.

Familiarity with the Questionnaire

If an interviewer is unfamiliar with the questionnaire, the study suffers and the respondent faces an unfair burden. The interview is likely to take more time than necessary and be unpleasant.

Moreover, the interviewer cannot acquire familiarity by skimming through the question-naire two or three times. He or she must study it carefully, question by question, and must practice reading it aloud.

Ultimately, the interviewer must be able to read the questionnaire items to respondents without error, without stumbling over words and phrases. A good model is the actor reading lines in a play or movie. The lines must be read as though they constituted a natural conversation, but that conversation must follow exactly the language set down in the questionnaire.

By the same token, the interviewer must be familiar with the specifications prepared in conjunction with the questionnaire. Inevitably some questions will not exactly fit a given respondent's situation, and the interviewer must determine how the question should be interpreted in that situation. The specifications provided to the interviewer should give adequate guidance in such cases, but the interviewer must know the organization and contents of the specifications well enough to refer to them efficiently. It would be better for the interviewer to leave a given question unanswered than to spend five minutes searching through the specifications for clarification or trying to interpret the relevant instructions.

Following Question Wording Exactly

The first part of this chapter discussed the sig-nificance of question wording for the responses obtained. A slight change in the wording of a given question may lead a respondent to answer "yes" rather than "no." It follows that interviewers must be instructed to follow the wording of questions exactly. Otherwise all the effort that the developers have put into carefully phrasing the questionnaire items to obtain the information they need and to ensure that respondents interpret items precisely as intended will be wasted.

While I hope the logic of this injunction is clear, it is not necessarily a closed discussion. For example, Giampietro Gobo (2006) argues that we might consider giving interviewers more latitude, suggesting that respondents sometimes make errors that may be apparent to the interviewer on the spot. As he notes, allowing the interviewer to intervene does increase the possibility that the interviewer will impact the data collected.

Recording Responses Exactly

Whenever the questionnaire contains open-ended questions (ones soliciting the respondent's own answers), the interviewer must record those answers exactly as given. No attempt should be made to summarize, paraphrase, or correct bad grammar.

This exactness is especially important because the interviewer will not know how the responses are to be coded. Indeed, the researchers them-selves may not know the coding until they've read a hundred or so responses. For example, the questionnaire might ask respondents how they feel about the traffic situation in their community. One respondent might answer that there are too many cars on the roads and that something should be done to limit their numbers. Another might say that more roads are needed. If the interviewer recorded these two responses with the same summary—"congested traffic"—the researchers would not be able to take advantage of the important differences in the original responses.

Sometimes, verbal responses are too inarticu-late or ambiguous to permit interpretation. How-ever, the interviewer may be able to understand the intent of the response through the respon-dent's gestures or tone. In such a situation, the interviewer should still record the exact verbal response but also add marginal comments giving both the interpretation and the reasons for arriving at it.

More generally, researchers can use any marginal comments explaining aspects of the response not conveyed in the verbal recording, such as the respondent's apparent anger, embar-rassment, uncertainty in answering, and so forth. In each case, however, the exact verbal response should also be recorded.

Probing for Responses

Sometimes respondents in an interview will give an inappropriate or incomplete answer. In such cases, a **probe**, or request for an elaboration, can

probe A technique employed in interviewing to solicit a more complete answer to a question. It is a nondirective phrase or question used to encourage a respondent to elaborate on an answer. Examples include "Anything more?" and "How is that?"

be useful. For example, a closed-ended question may present an attitudinal statement and ask the respondent to strongly agree, agree somewhat, disagree somewhat, or strongly disagree. The respondent, however, may reply: "I think that's true." The interviewer should follow this reply with "Would you say you strongly agree or agree somewhat?" If necessary, interviewers can explain that they must check one or the other of the categories provided. If the respondent adamantly refuses to choose, the interviewer should write in the exact response given by the respondent.

Probes are more frequently required in eliciting responses to open-ended than to closed-ended questions. For example, in response to a question about traffic conditions, the respondent might simply reply, "Pretty bad." The interviewer could obtain an elaboration on this response through a variety of probes. Sometimes the best probe is silence; if the interviewer sits quietly with pencil poised, the respondent will probably fill the pause with additional comments. (This technique is used effectively by newspaper reporters.) Appropriate verbal probes might be "How is that?" or "In what ways?" Perhaps the most generally useful probe is "Anything else?"

Often, interviewers need to probe for answers that will be sufficiently informative for analytical purposes. In every case, however, such probes must be completely neutral; they must not in any way affect the nature of the subsequent response. Whenever you anticipate that a given question may require probing for appropriate responses, you should provide one or more useful probes next to the question in the questionnaire. This practice has two important advantages. First, you'll have more time to devise the best, most neutral probes. Second, all interviewers will use the same probes whenever they're needed. Thus, even if the probe isn't perfectly neutral, all respondents will be presented with the same stimulus. This is the same logical guideline discussed for question wording. Although a question should not be loaded or biased, it's essential that every respondent be presented with the same question, even if it is biased.

Coordination and Control

Most interview surveys require the assistance of several interviewers. In large-scale surveys, interviewers are hired and paid for their work. Student researchers might find themselves recruiting friends to help them interview. Whenever more than one interviewer is involved in a survey, their efforts must be carefully controlled. This control has two aspects: training interviewers and supervising them after they begin work.

The interviewers' training session should begin with a description of what the study is all about. Even though the interviewers may be involved only in the data-collection phase of the project, it will be useful to them to understand what will be done with the interviews they conduct and what purpose will be served. Morale and motivation are usually lower when interviewers don't know what's going on.

The training on how to interview should begin with a discussion of general guidelines and procedures, such as those discussed earlier in this section. Then the whole group should go through the questionnaire together—question by question. Don't simply ask if anyone has any questions about the first page of the questionnaire. Read the first question aloud, explain the purpose of the question, and then entertain any questions or comments the interviewers may have. Once all their questions and comments have been handled, go on to the next question in the questionnaire.

It's always a good idea to prepare specifications to accompany an interview questionnaire. *Specifications* are explanatory and clarifying comments about handling difficult or confusing situations that may occur with regard to particular questions in the questionnaire. When drafting the questionnaire, try to think of all the problem cases that might arise—the bizarre circumstances that might make a question difficult to answer. The survey specifications should provide detailed guidelines on how to handle such situations. For example, even as simple a matter as age might present problems. Suppose a respondent says he or she will be 25 next week. The interviewer might not be sure whether to take the respondent's current age or the nearest one. The specifications for that question should explain what should be done. (Probably, you would specify that the age as of last birthday should be recorded in all cases.)

If you've prepared a set of specifications, review them with the interviewers when you

go over the individual questions in the questionnaire. Make sure your interviewers fully understand the specifications and the reasons for them as well as the questions themselves.

This portion of the interviewer training is likely to generate many troublesome questions from your interviewers. They'll ask, "What should I do if . . . ?" In such cases, avoid giving a quick, offhand answer. If you have specifications, show how the solution to the problem could be determined from the specifications. If you do not have specifications, show how the preferred handling of the situation fits within the general logic of the question and the purpose of the study. Giving unexplained answers to such questions will only confuse the interviewers and cause them to take their work less seriously. If you don't know the answer to such a question when it's asked, admit it and ask for some time to decide on the best answer. Then think out the situation carefully and be sure to give all the interviewers your answer, explaining your reasons.

Once you've gone through the whole questionnaire, conduct one or two demonstration interviews in front of everyone. Preferably, you should interview someone other than one of the interviewers. Realize that your interview will be a model for those you're training, so make it good. It would be best, moreover, if the demonstration interview were done as realistically as possible. Don't pause during the demonstration to point out how you've handled a complicated situation: Handle it, and then explain later. It's irrelevant if the person you're interviewing gives real answers or takes on some hypothetical identity for the purpose, as long as the answers are consistent.

After the demonstration interviews, pair off your interviewers and have them practice on each other. When they've completed the questionnaire, have them reverse roles and do it again. Interviewing is the best training for interviewing. As your interviewers practice on each other, wander around, listening in on the practice so you'll know how well they're doing. Once the practice is completed, the whole group should discuss their experiences and ask any other questions they may have.

The final stage of the training for interviewers should involve some "real" interviews. Have them conduct some interviews under the actual conditions that will pertain to the final survey. You may want to assign them people to interview, or perhaps they may be allowed to pick people themselves. Don't have them practice on people you've selected in your sample, however. After each interviewer has completed three to five interviews, have him or her check back with you. Look over the completed questionnaires for any evidence of misunderstanding. Again, answer any questions that the interviewers have. Once you're convinced that a given interviewer knows what to do, assign some actual interviews, using the sample you've selected for the study.

It's essential to continue supervising the work of interviewers over the course of the study. You should check in with them after they conduct no more than 20 or 30 interviews. You might assign 20 interviews, have the interviewer bring back those questionnaires when they're completed, look them over, and assign another 20 or so. Although this may seem overly cautious, you must continually protect yourself against misunderstandings that may not be evident early in the study. Moreover, Kristen Olson and Andy Peytchev (2007) have discovered that interviewers' behavior continues to change over the course of a survey project. For example, as time goes on, interviewers speed through the interview more quickly and are more likely to judge respondents as uninterested in it.

If you're the only interviewer in your study, these comments may not seem relevant. However, it would be wise, for example, to prepare specifications for potentially troublesome questions in your questionnaire. Otherwise, you run the risk of making ad hoc decisions, during the course of the study, that you'll later regret or forget. Also, the emphasis on practice applies equally to the one-person project and to the complex funded survey with a large interviewing staff.

Telephone Surveys

For years telephone surveys had a rather bad reputation among professional researchers. By definition, telephone surveys are limited to people who have telephones. Years ago, this method produced a substantial social-class bias by excluding poor people from the surveys. This was vividly demonstrated by the *Literary Digest*

fiasco of 1936. Recall that, even though voters were contacted by mail, the sample was partially selected from telephone subscribers, who were hardly typical in a nation just recovering from the Great Depression. As we saw in Chapter 7, virtually all American households now have telephones, so the earlier form of class bias has substantially diminished.

Telephone surveys offer many advantages that underlie the popularity of this method. Probably the greatest returns are in money and time, in that order. To conduct a face-to-face, household interview, you may drive several miles to a respondent's home, find no one there, return to the research office, and drive back the next day—possibly finding no one there again. It's cheaper and quicker to let your fingers make the trips.

Interviewing by telephone, you can dress any way you please without affecting the answers respondents give. And sometimes respondents will be more honest in giving socially disapproved answers if they don't have to look you in the eye. Similarly, it may be possible to probe into more-sensitive areas, though this isn't necessarily the case. People are, to some extent, more suspicious when they can't see the person asking them questions.

Interviewers can communicate a lot about themselves over the phone, however, even though they can't be seen. For example, researchers worry about the impact of an interviewer's name (particularly if ethnicity is relevant to the study) and debate the ethics of having all interviewers use bland "stage names" such as Smith or Jones. (Female interviewers sometimes ask permission to do this, to avoid subsequent harassment from men they interview.)

Telephone surveys can allow greater control over data collection if several interviewers are engaged in the project. If all the interviewers are calling from the research office, they can get clarification from the person in charge whenever problems occur, as they inevitably do. Alone in

the boondocks, an interviewer may have to wing it between weekly visits with the interviewing supervisor.

Telephone interviewing presents its own problems, however. For example, the method is hampered by the proliferation of bogus "surveys" that are actually sales campaigns disguised as research. If you have any questions about any such call you receive, by the way, ask the interviewer directly whether you've been selected for a survey only or if a sales "opportunity" is involved. It's also a good idea, if you have any doubts, to get the interviewer's name, phone number, and company. Hang up if the caller refuses to provide any of these.

For the researcher, the ease with which people can hang up is another shortcoming of telephone surveys. Once you've been let inside someone's home for an interview, the respondent is unlikely to order you out of the house in mid-interview. It's much easier to terminate a telephone interview abruptly, saying something like, "Whoops! Someone's at the door. I gotta go." or "Omigod! The neighbors are setting my car on fire!" (That sort of evasion is much harder to fake when the interviewer is sitting in your living room.)

Research has shown that several factors, including voice mail and answering machines, have reduced response rates in telephone surveys. Peter Tuckel and Harry O'Neill (2002) and others have examined the impact of such factors as Caller ID, answering machines, and telemarketing. All these constitute difficulties modern survey researchers must deal with.

Computer-Assisted Telephone Interviewing (CATI)

In Chapter 14, we'll see some of the ways computers have influenced the conduct of social research—particularly data processing and analysis. Computers are also changing the nature of telephone interviewing. One innovation is **computer-assisted telephone interviewing (CATI)**. This method is increasingly used by academic, government, and commercial survey researchers. Though there are variations in practice, here's what CATI can look like.

Imagine an interviewer wearing a telephone headset, sitting in front of a computer terminal

computer-assisted telephone interviewing (CATI) A data-collection technique in which a telephone-survey questionnaire is stored in a computer, permitting the interviewer to read the questions from the monitor and enter the answers on the computer keyboard.

and its video screen. The central computer selects a telephone number at random and dials it. On the video screen is an introduction ("Hello, my name is . . .") and the first question to be asked ("Could you tell me how many people live at this address?").

When the respondent answers the phone, the interviewer says hello, introduces the study, and asks the first question displayed on the screen. When the respondent answers the question, the interviewer types that answer into the computer terminal—either the verbatim response to an open-ended question or the code category for the appropriate answer to a closed-ended question. The answer is immediately stored in the computer. The second question appears on the video screen, is asked, and the answer is entered into the computer. Thus, the interview continues.

In addition to the obvious advantages in terms of data collection, CATI automatically prepares the data for analysis; in fact, the researcher can begin analyzing the data before the interviewing is complete, thereby gaining an advanced view of how the analysis will turn out.

It is also possible to go a step further than computer-*assisted* interviews. With the innovation of so-called robo-polls, the entire interview is conducted by a programmed recording that can interpret the spoken answers of respondents. This discussion may remind you of the robo-calls in which a recorded voice presents a political or commercial message once you answer your phone. Robo-polls go a step further through the use of *Interactive Voice Recognition (IVR)*. The computer is programmed to interpret the respondent's answers, record them, and determine how to continue the interview appropriately.

Clearly this method is cost-effective by cutting out the labor cost of hiring human beings as interviewers. It has been viewed with suspicion and/or derision by some survey researchers, but in its evaluation of the 2008 primary polling, the American Association of Public Opinion Research (AAPOR) reported no difference in the accuracy of results produced by CATI or IVR (AAPOR 2009).

During the 2010 midterm election campaigns, survey-watcher Nate Silver (2010b) found that robo-polls tended to produce results slightly more favorable to Republicans than did conventional methods. Silver also found that robo-polls might produce different answers to sensitive questions. He looked at California's Proposition 19, which would have legalized and taxed the personal use of marijuana. Silver found:

> The methodologies split in the support they show for the initiative. The three automated surveys all have Prop 19 passing by a double-digit margin. The human-operator polls, meanwhile, each show it trailing narrowly.
>
> *(Silver: 2010a)*

Ultimately, Proposition 19 failed by a two-to-one margin. The next edition of this textbook may revise the discussion of robo-polls, though it is not clear now what the fate of this technique will be.

Response Rates in Interview Surveys

Earlier in this chapter we looked at the issue of response rates in mail surveys, and this is an equally important issue for interview surveys. In Chapter 7, when we discussed formulas for calculating sampling error to determine the accuracy of survey estimates, the implicit assumption was that everyone selected in a sample would participate—which is almost never the case. Lacking perfection, researchers must maximize participation by those selected. Although interview surveys tend to produce higher response rates than do mail surveys, interview success has recently declined.

By analyzing response-rate trends in the University of Michigan's Survey of Consumer Attitudes, Richard Curtin, Stanley Presser, and Eleanor Singer (2005) have sketched a pattern of general decline over recent years. Between 1979 and 1996, the response rate in this telephone survey dropped from 72 to 60 percent, representing an average annual decline of three-quarters of a percent. Since 1996, the rate of decline has doubled. The increased non-responses reflected both refusals and those who the interviewers were unable to contact.

By contrast, the General Social Survey, using personal interviews, experienced response rates between 73.5 and 82.4 percent in the years from 1975 to 1998. In the 2000 and 2002 surveys, however, the GSS completion rate was 70 percent.

Their decline came primarily from refusals rather than being unable to contact respondents, because household interviews produce higher rates of contact than telephone surveys do.

In recent years, both household and telephone surveys have experienced a decline in response rates. A special issue of the *Public Opinion Quarterly* (2006) was devoted entirely to analyzing the many dimensions of the decline in response rates in household surveys. As the analyses show, lower response rates do not necessarily produce inaccurate estimates of the population being studied, but the variations on this issue defy a simple summary.

Former director of the U.S. Census, Robert Groves (2011: 866) detailed some of the factors complicating modern survey research.

> Walled subdivisions, locked apartment buildings, telephone answering machines, telephone caller ID, and a host of other access impediments for survey researchers grew in this era. Response rates continued to deteriorate. Those household surveys devoted to high response rates experienced continuous inflation of costs due to increased effort to contact and interview the public. Face-to-face interviews continued to decline in volume, often limited to the first wave of longitudinal surveys.

Many researchers believe that the widespread growth of telemarketing has been a big part of the problems experienced by legitimate telephone surveys, and there are hopes that the state and national "do not call" lists may ease that problem. Further, as we've seen, other factors such as answering machines and voicemail also contribute to these problems (Tuckel and O'Neill 2002). Response rate is likely to remain an issue of high concern in survey research.

As a consumer of social research, you should be wary of "surveys" whose apparent purpose is to raise money for the sponsor. This practice had been common in mail surveys, and soon expanded to the realm of "fax surveys," evidenced by a fax entitled "Should Hand Guns Be Outlawed?" Two fax numbers were provided for expressing either a "Yes" or "No" opinion. The smaller print noted, "Calls to these numbers cost $2.95 per minute, a small price for greater democracy. Calls take approx. 1 or 2 minutes."

You can imagine where the $2.95 went. Undoubtedly, you can give your own examples of similar e-mail "surveys."

Online Surveys

An increasingly popular method of survey research involves the use of the Internet, one of the most far-reaching developments of the late twentieth century. Mick Couper and Peter Miller (2008) give an excellent introduction to the timeline of this new face of social research.

> Despite their relatively short history, Web surveys have already had a profound effect on survey research. The first graphic browser (NCSA Mosaic) was released in 1992, with Netscape Navigator following in 1994 and Internet Explorer in 1995. The first published papers on Web surveys appeared in 1996. Since then, there has been a virtual explosion of interest in the Internet as a tool for survey data collection.
>
> *(2008: 831)*

Three years later, Couper (2011) reflected on the probable role of online surveys in the future of social research.

> The newer modes have tended to supplement rather than replace existing modes, in part because even though they address some problems (e.g., improvements in measurement, reductions in cost), they may not solve others (e.g., coverage, nonresponse). In other words, there is no one mode that can be all things to all research questions. Multiple modes, and mixes of mode, will continue to be a fact of life for survey research for the foreseeable future.
>
> *(2011: 901)*

While this section will examine various aspects of online survey research, you should be forewarned that this technique is developing so quickly that new innovations will surely have arisen by the time this book reaches your hands. To stay abreast of these developments, your best single source is the American Association for Public Opinion Research (AAPOR) and two key publications: *Public Opinion Quarterly* (POQ) and the online journal *Survey Practice.* Although

neither of these is solely dedicated to online research, an increasing percentage of their articles addresses that topic. University survey research offices such as those at the University of Michigan, NORC at the University of Chicago, and many other institutions around the globe are very active in developing this new technique. Similarly, commercial research firms such as Pew, Harris, Nielsen, and others are equally involved.

As we saw in Chapter 7 on sampling, one immediate objection that many social researchers make to online surveys concerns representativeness: Will the people who can be surveyed online be representative of meaningful populations, such as all U.S. adults, all voters, and so on? This was the criticism raised previously with regard to surveys via fax or by telephone interviewers.

Early in the development of online surveys, Camilo Wilson (1999), founder of Cogix, pointed out that some respondent populations are ideally suited to this technique: specifically, those who visit a particular website. For example, Wilson indicates that market research for online companies should be conducted online, and his firm has developed software called ViewsFlash for precisely that purpose. Although website surveys could easily collect data from all who visit a particular site, Wilson suggests that survey-sampling techniques can provide sufficient consumer data without irritating thousands or millions of potential customers. As we saw in Chapter 7, much methodological research is being devoted to ways of achieving representative sampling of general populations with online surveys.

Let's turn now to some of the other methodological aspects of online surveys that are currently being examined and experimented with.*

Online Devices

At the outset, online surveys were aimed at users of personal computers, most typically desktop models. As the distinction between desktop and

*In beginning this section of the chapter, I want to acknowledge Michael Link of the Nielsen Company, for his excellent, online seminar, "Leveraging New Technologies," conducted as part of AAPOR's Webinar Series on December 5, 2012. While I have not quoted directly from the seminar, I have benefited greatly from the overview and detailing of variations it provided.

laptop computer capabilities narrowed, both devices were considered proper ways of participating in online surveys. Notice, however, that the growing use of laptop computers for this purpose broadened the variety of environments in which respondents might participate. This was only the beginning, however.

When I attended the first meeting of the Chinese Survey Research Association in Shanghai in 2010, I was struck by the vitality of the researchers reporting on their studies in a country where sociology had been removed from universities from 1949 to 1979. Most of the articles I looked at were in Chinese, which was a problem for me. However, many articles included photographs to illustrate some of the new techniques being used, and I was struck by the number of smartphones and other mobile devices pictured. This interest is hardly limited to Chinese research.

Tablets and smartphones have been rapidly gaining in computing power and are increasingly being used as vehicles for completing online surveys. Respondents have inadvertently compelled researchers to develop survey formats that were compatible with mobile devices: As respondents attempted, sometimes unsuccessfully, to use smartphones and digital tablets to complete questionnaires designed for desktop computers, survey researchers realized the need and potential for adapting their questionnaires to the range of devices that might be used by respondents. Screen size, of course, is a major concern, but so are the varied navigation systems used by different devices.

Researchers are also learning that they must accommodate respondents' device preferences. For example, Morgan M. Millar and Don A. Dillman (2012) conducted an experiment in which they attempted to encourage respondents to participate in a survey using their smartphones while still allowing the use of other devices such as tablets or laptops. The researchers reported only a slight increase in smartphone usage by respondents who were urged to use the device, compared with those who were given no encouragement.

This line of methodological research will continue, but consider this: We will surely see the development of new devices, some we can't currently imagine, that will have to be accommodated in the future.

Electronic Instrument Design

Over the years, members of industrialized nations have become familiar with the format and process of self-administered questionnaires, but, as just mentioned, the web presents a new challenge for many. Leah Christian, Don Dillman, and Jolene Smyth provide a wealth of guidance on the formatting of web surveys. Their aim is, as their article title suggests, "helping respondents get it right the first time" (2007).

The initial temptation, of course, is to simply import the digital file for the mail questionnaire into a web survey framework. However, there are two problems with this. First, the mail format doesn't necessarily fit on a computer screen, let alone onto that of a tablet or smartphone. On the other hand, the e-devices offer possibilities unattainable with words on paper. I am unable to list those possibilities for you now, because they are still being developed, but I can connect you with some of the options and challenges currently underway or on the radar.

For example, researchers like Roger Tourangeau, Mick P. Couper, and Frederick G. Conrad (2013) were concerned about whether the placement of answers in a list would affect respondents' choices. Their conclusion, based on the review of several studies, is that "up means good." When several opinion choices are arranged vertically, respondents are more likely to select the topmost choice.

Jason Husser and Kenneth Fernandez (2013) examined whether it was better to have an online respondent enter numerical answers by clicking the answer, typing it, or drag along a scale to indicate the answer. With a limited number of responses, clicking radio buttons was fastest, but a long list of possible answers makes dragging the sliding scale more practical.

Those regularly using the Internet are familiar with emoticons such as the "smiley face." While these graphics could be printed in a mail questionnaire, they seem more at home online. Matthias Emde and Marek Fuchs (2012) undertook an experiment to determine the possibility of using a range of faces (sad to happy) in place of radio buttons labeled from bad to good. They concluded that this format change did not affect responses. Thus, these types of formatting options may be

chosen on purely aesthetic grounds. There is no reason not to make surveys appealing.

Malakhoff and Jans (2011) explore some of the more advanced possibilities for online survey research. While the survey interview involves a person showing up on your doorstep or a voice coming over your phone, they suggest that an animated avatar might be used to conduct an online interview, and they have begun experimenting with gender and other differences for the animated interviewer. The avatar interviewer can be programmed to change facial expressions based on the respondent's answers. Going one step (or several) further, it would be possible to use the respondents' webcams to monitor their facial expressions and log that data along with the answers provided verbally.

The relative youth of online surveys makes them a fertile ground for innovation and experimentation. For example, survey researchers have often worried that respondents to self-administered questionnaires may spend more of their attention on the first responses in a list, skipping quickly over those farther down. To test this possibility, Mirta Galesic and colleagues (2008) employed a special eye-tracking computer monitor that unobtrusively followed respondents' eye movements as they completed an online survey. The result: Respondents did, in fact, spend more time on the early choices, sometimes failing to read the whole list before clicking their choice on the screen. We may expect to see more such experimentation in the future.

Improving Response Rates

Online surveys appear to have response rates approximately comparable to mail surveys, according to a large-scale study of Michigan State University students (Kaplowitz, Hadlock, and Levine 2004), especially when the online survey is accompanied by a postcard reminder encouraging respondents to participate. While producing a comparable response rate, the cost of the online survey is substantially less than that of a conventional mail survey. The cost of paper, printing, and postage alone can constitute a large expense.

In another study of ways to improve response rates in online surveys, Stephen Porter and Michael Whitcomb (2003) found that some

Tips and Tools

Conducting an Online Survey

If you're interested in testing the waters of online surveys, Survey Monkey™ may give you one opportunity to try your hand at this emerging technique. At this writing, you can sign up to experiment with a limited version of the online survey program at no charge. Visit www.surveymonkey.com/ and follow the instructions on the website.

You will be shown how to construct the questionnaire and enter the e-mail addresses of those you wish to survey. Once the responses come in from your subjects, you will be able to conduct an analysis of your data.

You can use Survey Monkey with a limited number of friends to sharpen your survey research skills, and/or you can use it for a full-blown, professional study. In fact, it is sometimes used by professional researchers and research associations.

of the techniques effective in mail surveys, such as personalizing the appeal or varying the apparent status of the researcher, had little or no impact in the new medium. At the same time, specifying that the respondents had been specially selected for the survey and setting a deadline for participation did increase response rates. The years ahead will see many experiments aimed at improving the effectiveness of online surveys.

You are reading this discussion at an exciting time, when online survey methodology is evolving. For example, in an effort to increase response rates for web surveys, Morgan Millar and Don Dillman (2012) achieved modest increases by sending respondents an e-mail reminder to participate in the survey. Because a large percentage of cell phone owners have smartphones, they were offered the opportunity to complete the survey on those devices instead of going to a computer. As the authors point out, further experimentation with e-mail reminders will require tailoring survey formats to accommodate smartphones as discussed earlier.

For now, Mick P. Couper's *Designing Effective Web Surveys* (2008) offers a comprehensive guide to this new technique, based on what we have learned about it to date. If you are interested in experimenting with web surveys on your own, see the Tips and Tools box, "Conducting an Online Survey."

Mixed-Mode Surveys

In Chapter Four, I introduced the idea of mixed modes, indicating that different research techniques could be combined in a given study: such

as a survey, combined by a review of existing data and in-depth field observations and interviews. Although researchers have sometimes combined face-to-face, mail, and telephone surveys, the advent of online surveys has increased attention to the potential of combining survey techniques.

As Don Dillman (2012) points out, the logistical advantages of online surveys are somewhat offset by the difficulty of getting representative samples. Thus, researchers sometimes use an address-based sampling as the basis for a mail survey, which invites recipients to respond online if that's convenient for them, or by mail if it is not.

As Edith de Leeuw (2010) points out, this is not a new idea.

> Already in 1788, Sir John Sinclair used a mixed-mode approach. Lacking funds for a full statistical census, Sinclair used a cost-effective mail survey among ministers of all parishes in the Church of Scotland. To achieve a high response Sinclair also used follow-up letters and finally "statistical missionaries," who personally visited the late responders to hurry ministerial replies.

This combination of survey techniques evidently produced a 100 percent completion rate.

The special advantages of Internet surveys (mass scale and cost) have added new impetus for combining survey modes. In addition to sampling issues, survey researchers are also attentive to response effects that may be caused by the different modes. That is, whether people would answer a given question the same online as in a mail questionnaire or a telephone interview. Initial

studies suggest relatively small effects (De Leeuw and Hox 2012), but this will be a subject of methodological research for years to come.

Comparison of the Different Survey Methods

Now that we've seen several ways to collect survey data, let's take a moment to compare them directly.

Self-administered questionnaires are generally cheaper and quicker than face-to-face interview surveys. These considerations are likely to be important for an unfunded student wishing to undertake a survey for a term paper or thesis. Moreover, if you use the self-administered mail format, it costs no more to conduct a national survey than a local one of the same sample size. In contrast, a national interview survey utilizing face-to-face contacts would cost far more than a local one. Also, mail surveys typically require a small staff: You could conduct a reasonable mail survey by yourself, although you shouldn't underestimate the work involved. Further, respondents are sometimes reluctant to report controversial or deviant attitudes or behaviors in interviews but are willing to respond to an anonymous self-administered questionnaire.

Interview surveys also offer many advantages. For example, they generally produce fewer incomplete questionnaires. Although respondents may skip questions in a self-administered questionnaire, interviewers are trained not to do so. In CATI surveys, the computer offers a further check on this. Interview surveys, moreover, have typically achieved higher completion rates than self-administered questionnaires have.

Although self-administered questionnaires may be more effective for sensitive issues, interview surveys are definitely more effective for complicated ones. Prime examples include the enumeration of household members and the determination of whether a given address corresponds to more than one housing unit. Although the concept of housing unit has been refined and standardized by the Census Bureau and interviewers can be trained to deal with the concept, it's extremely difficult to communicate this idea in a self-administered questionnaire.

This advantage of interview surveys pertains generally to all complicated contingency questions.

With interviews, you can conduct a survey based on a sample of addresses or phone numbers rather than on names. An interviewer can arrive at an assigned address or call the assigned number, introduce the survey, and even—following instructions—choose the appropriate person at that address to respond to the survey. In contrast, self-administered questionnaires addressed to "occupant" receive a notoriously low response.

Finally, as we've seen, interviewers questioning respondents face-to-face can make important observations aside from responses to questions asked in the interview. In a household interview, they may note the characteristics of the neighborhood, the dwelling unit, and so forth. They can also note characteristics of the respondents or the quality of their interaction with the respondents—whether the respondent had difficulty communicating, was hostile, seemed to be lying, and so on. A student using this textbook recently pointed out another advantage of face-to-face interviews. In his country, where literacy rates are relatively low in some areas, people would not be able to read a self-administered questionnaire and record their answers—but they could be interviewed.

The chief advantages of telephone surveys over those conducted face-to-face center primarily on time and money. Telephone interviews are much cheaper and can be mounted and executed quickly. Also, interviewers are safer when interviewing people living in high-crime areas. Moreover, the impact of the interviewers on responses is somewhat lessened when the respondents can't see them. As only one indicator of the popularity of telephone interviewing, when Johnny Blair and his colleagues (1995) compiled a bibliography on sample designs for telephone interviews, they listed over 200 items.

Online surveys have many of the strengths and weaknesses of mail surveys. Once the available software has been further developed, they will likely be substantially cheaper. An important weakness, however, lies in the difficulty of assuring that respondents to an online survey will be representative of some more-general population.

Martyn Denscombe (2009) used matched samples of students to test the non-response rates produced by conventional, paper questionnaires with those administered online. (Students did not get to choose the method but were randomly assigned.) Overall, the online surveys produced somewhat lower non-response rates, and this difference was more pronounced for open-ended questions.

Online surveys are particularly appropriate for certain targeted groups, and research specifically based on web participation. An online survey would be perfect for studying the feelings of those people who have purchased items from Seller #12345 on eBay, for example. This advantage may become more significant if and when our lives become increasingly organized around our web participation.

As respondents become more accustomed to online surveys, it may ease some of the problems that have plagued telephone surveys, such as allowing for longer and more-complex surveys. Online respondents, like those completing mail questionnaires will have more time to reflect on their responses. In addition, online surveys may lend themselves to experimental designs more easily than other methods. As took place with earlier survey techniques, online survey methodology will continue to evolve as it is increasingly utilized by researchers.

With the growth of online surveys, we have seen an increased interest in and use of *paradata*, a wealth of data generated by computer in the course of a survey. How long did a respondent take before answering each question? Did men or women take longer to answer a particular question? Did conservative or liberal responses come more quickly? Already such data are being used for studies of survey methodology, but they also can provide data useful to understanding human behavior, as social scientists are wont to do.

Clearly, each survey method has its place in social research. Ultimately, you must balance the advantages and disadvantages of the different methods in relation to your research needs and your resources. As we have just seen, researchers sometimes employ mixed-mode surveys in the same study, combining more than one of the techniques we've examined, such as mail and interview. While this option has been employed

for some time, Edith D. de Leeuw (2010) updated the discussion by bringing online surveys into the mix.

Strengths and Weaknesses of Survey Research

Regardless of the specific method used, surveys—like other modes of observation in social research—have special strengths and weaknesses. You should keep these in mind when determining whether a survey is appropriate for your research goals.

Surveys are particularly useful in describing the characteristics of a large population. A carefully selected probability sample in combination with a standardized questionnaire offers the possibility of making refined descriptive assertions about a student body, a city, a nation, or any other large population. Surveys determine unemployment rates, voting intentions, and so forth with uncanny accuracy. Although the examination of official documents—such as marriage, birth, or death records—can provide equal accuracy for a few topics, no other method of observation can provide this general capability.

Surveys—especially self-administered ones—make large samples feasible. Surveys of 2,000 respondents are not unusual. A large number of cases is very important for both descriptive and explanatory analyses, especially wherever several variables are to be analyzed simultaneously.

In one sense, surveys are flexible. Many questions can be asked on a given topic, giving you considerable flexibility in your analyses. Whereas an experimental design may require you to commit yourself in advance to a particular operational definition of a concept, surveys let you develop operational definitions from actual observations.

Finally, standardized questionnaires have an important strength in regard to measurement generally. Earlier chapters have discussed the ambiguous nature of most concepts: They have no ultimately real meanings. One person's religiosity is quite different from another's. Although you must be able to define concepts in those ways most relevant to your research goals, you may not find it easy to apply the same definitions uniformly to all subjects. The survey researcher

is bound to this requirement by having to ask exactly the same questions of all subjects and having to impute the same intent to all respondents giving a particular response.

Survey research also has several weaknesses. First, the requirement of standardization often seems to result in the fitting of round pegs into square holes. Standardized questionnaire items often represent the least common denominator in assessing people's attitudes, orientations, circumstances, and experiences. By designing questions that will be at least minimally appropriate to all respondents, you may miss what is most appropriate to many respondents. In this sense, surveys often appear superficial in their coverage of complex topics. Although this problem can be partly offset by sophisticated analyses, it is inherent in survey research.

Similarly, survey research can seldom deal with the context of social life. Although questionnaires can provide information in this area, the survey researcher rarely develops the feel for the total life situation in which respondents are thinking and acting that, say, the participant observer can (see Chapter 10).

In many ways, surveys are inflexible. Studies involving direct observation can be modified as field conditions warrant, but surveys typically require that an initial study design remain unchanged throughout. As a field researcher, for example, you can become aware of an important new variable operating in the phenomenon you're studying and begin making careful observations of it. The survey researcher would probably be unaware of the new variable's importance and could do nothing about it in any event.

Finally, surveys are subject to the artificiality mentioned earlier in connection with experiments. Finding out that a person gives conservative answers in a questionnaire does not necessarily mean the person is conservative; finding out that a person gives prejudiced answers in a questionnaire does not necessarily mean the person is prejudiced. This shortcoming is especially salient in the realm of action. Surveys cannot measure social action; they can only collect self-reports of recalled past action or of prospective or hypothetical action.

The problem of artificiality has two aspects. First, the topic of study may not be amenable to measurement through questionnaires. Second, the act of studying that topic—an attitude, for example—may affect it. A survey respondent may have given no thought to whether the governor should be impeached until asked for his or her opinion by an interviewer. He or she may, at that point, form an opinion on the matter.

Survey research is generally weak on validity and strong on reliability. In comparison with field research, for example, the artificiality of the survey format puts a strain on validity. As an illustration, people's opinions on issues seldom take the form of strongly agreeing, agreeing, disagreeing, or strongly disagreeing with a specific statement. Their survey responses in such cases must be regarded as approximate indicators of what the researchers had in mind when they framed the questions. This comment, however, needs to be held in the context of earlier discussions of the ambiguity of validity itself. To say something is a valid or an invalid measure assumes the existence of a "real" definition of what's being measured, and many scholars now reject that assumption.

Reliability is a clearer matter. By presenting all subjects with a standardized stimulus, survey research goes a long way toward eliminating unreliability in observations made by the researcher. Moreover, careful wording of the questions can also significantly reduce the subject's own unreliability.

As with all methods of observation, a full awareness of the inherent or probable weaknesses of survey research can partially resolve them in some cases. Ultimately, though, researchers are on the safest ground when they can employ several research methods in studying a given topic.

Secondary Analysis

As a mode of observation, survey research involves the following steps: (1) questionnaire construction, (2) sample selection, and (3) data collection, through either interviewing or self-administered questionnaires. As you've gathered, surveys are usually major undertakings. It's not unusual for a large-scale survey to take several months or even more than a year to progress from conceptualization to data in hand. (Smaller-scale surveys can, of course, be done

more quickly.) Through a method called secondary analysis, however, researchers can pursue their particular social research interests—analyzing survey data from, say, a national sample of 2,000 respondents—while avoiding the enormous expenditure of time and money such a survey entails.

Secondary analysis is a form of research in which the data collected and processed by one researcher are reanalyzed—often for a different purpose—by another. Beginning in the 1960s, survey researchers became aware of the potential value that lay in archiving survey data for analysis by scholars who had nothing to do with the survey design and data collection. Even when one researcher had conducted a survey and analyzed the data, those same data could be further analyzed by others who had slightly different interests. Thus, if you were interested in the relationship between political views and attitudes toward gender equality, you could examine that research question through the analysis of any data set that happened to contain questions relating to those two variables.

The initial data archives were very much like book libraries, with a couple of differences. First, instead of books, the data archives contained data sets: first as punched cards, then as magnetic tapes. Today they're typically contained on computer hard drives, portable electronic storage devices, or online servers. Second, whereas you're expected to return books to a conventional library, you can keep the data obtained from a data archive.

The best-known current example of secondary analysis is the General Social Survey (GSS). The National Opinion Research Center (NORC) at the University of Chicago conducts this major national survey, currently every other year, to collect data on a large number of social science variables. These surveys are conducted precisely for the purpose of making data available to scholars at little or no cost and are supported by a combination of private and government funding. Recall that the GSS was created by James A. Davis in 1972; it is currently directed by Davis, Tom W. Smith, and Peter V. Marsden. Their considerable ongoing efforts make an unusual contribution to social science research and to education in social science.

Numerous other resources are available for identifying and acquiring survey data for secondary analysis. The Roper Center for Public Opinion Research at the University of Connecticut is one excellent resource. The center also publishes the journal *Public Perspective,* which is focused on public opinion polling.

Because secondary analysis has typically involved obtaining a data set and undertaking an extensive analysis, I would like you to consider another approach as well. Often you can do limited analyses by investing just a little time. Let's say you're writing a term paper about the impact of religion in contemporary American life. You want to comment on the role of the Roman Catholic Church in the debate over abortion. Although you might get away with an off-hand, unsubstantiated assertion, imagine how much more powerful your paper would be if you supported your position with additional information. Follow the steps in Figure 9-7 to learn how to access data relevant to this research topic.

1. Go to the SDA analysis site at http://sda .berkeley.edu/sdaweb/analysis/?dataset=gss12, which was introduced in Chapter 1.

2. In the codebook listing on the left of the figure, locate the survey items dealing with abortion—by selecting the appropriate entry under "Controversial Social Issues."

3. For purposes of this illustration, let's see how members of the different religious groups responded in regard to women being allowed to choose an abortion "for any reason."

4. Type the name of this item—ABANY—where I have entered it in Figure 9-7.

5. Locate the variable label for Religious Affiliation in the column to the left, and enter RELIG where I have entered it in Figure 9-7. And to see current opinions on this topic, specify the year 2012 as I have done in the Figure.

6. Click the button labeled "Run the Table" and you should be rewarded with the table shown in Figure 9-8.

secondary analysis A form of research in which the data collected and processed by one researcher are reanalyzed—often for a different purpose—by another. This is especially appropriate in the case of survey data. Data archives are repositories or libraries for the storage and distribution of data for secondary analysis.

FIGURE 9-7

Requesting an Analysis of GSS Data.

Source: SDA at http://sda.berkeley.edu/sdaweb/analysis/?dataset=gss12.

The results of your analysis, shown in Figure 9-8, may surprise you. Whereas Catholics are less supportive of abortion (38.1 percent) than Jews (90 percent) and those with no religion (63.3 percent), they are slightly *more* supportive than American Protestants (37.1 percent).

Imagine a term paper that says, "Whereas the Roman Catholic Church has taken a strong, official position on abortion, many Catholics do not necessarily agree, as shown in Table . . ." Moreover, this might be just the beginning of

an analysis that looks a bit more deeply into the matter, as will be described in Chapter 14, where we discuss quantitative analysis.

The key advantage of secondary analysis is that it's cheaper and faster than doing original surveys, and, depending on who did the original survey, you may benefit from the work of topflight professionals. The ease of secondary analysis has also enhanced the possibility of meta-analysis, in which a researcher brings together a body of past research on a particular topic. To gain confidence in your understanding

SDA 4.0: Tables

General Social Survey Cumulative Datafile 1972-2012

Apr 23, 2014 (Wed 12:17 PM PDT)

Variables					
Role	Name	Label	Range	MD	Dataset
Row	ABANY	ABORTION IF WOMAN WANTS FOR ANY REASON	1-2	0,8,9	1
Column	RELIG	RS RELIGIOUS PREFERENCE	1-13	0,98,99	1
Weight	COMPWT	Composite weight = WTSSALL * OVERSAMP * FORMWT	1925-11.1207		1
Filter	YEAR(2012)	GSS YEAR FOR THIS RESPONDENT	1972-2012		1

Frequency Distribution														
Cells contain: -Column percent -Weighted N							RELIG							
		1 PROTESTANT	2 CATHOLIC	3 JEWISH	4 NONE	5 OTHER	6 BUDDHISM	7 HINDUISM	8 OTHER EASTERN	9 MOSLEM/ISLAM	10 ORTHODOX-CHRISTIAN	11 CHRISTIAN	12 NATIVE AMERICAN	13 INTER-NONDENOMINATIONAL	ROW TOTAL
ABANY	1: YES	37.1 / 206.7	38.1 / 120.8	90.0 / 15.3	63.3 / 143.4	62.6 / 11.6	75.0 / 1.2	71.5 / 5.1	85.3 / 5.0	8.7 / 1.6	28.5 / .8	32.2 / 24.5	0 / .0	13.5 / .4	42.8 / 536.5
	2: NO	62.9 / 349.9	61.9 / 196.0	10.0 / 1.7	36.7 / 83.2	37.4 / 6.9	25.0 / .4	28.5 / 2.0	14.7 / .9	91.3 / 17.2	71.5 / 2.1	67.8 / 51.4	100.0 / 2.5	86.5 / 2.6	57.2 / 716.7
	COL TOTAL	100.0 / 556.6	100.0 / 316.7	100.0 / 16.9	100.0 / 226.6	100.0 / 18.6	100.0 / 1.6	100.0 / 7.2	100.0 / 5.9	100.0 / 18.8	100.0 / 2.9	100.0 / 75.9	100.0 / 2.5	100.0 / 3.0	100.0 / 1,253.2

Color coding: <-2.0 <-1.0 <0.0 >0.0 >1.0 >2.0 Z

N in each cell: Smaller than expected Larger than expected

of the relationship between religion and abortion, for example, you could go beyond the GSS to analyze similar data collected in dozens or even hundreds of other studies.

There are disadvantages inherent in secondary analysis, however. The key problem involves the recurrent question of validity. When one researcher collects data for one particular purpose, you have no assurance that those data will be appropriate for your research interests. Typically, you'll find that the original researcher asked a question that "comes close" to measuring what you're interested in, but you'll wish the question had been asked just a little differently—or that another, related question had also been asked. For example, you may want to study how religious various people are and the survey data available to you only asked about attendance at worship services. Your quandary, then, is whether the question that was asked provides a valid measure of the variable you want to analyze. Nevertheless, secondary analysis can be immensely useful. Moreover, it illustrates once again the range of possibilities available in finding the answers to questions about social life. Although no single method unlocks all puzzles, there is no limit to the ways you can find out about things. And when you zero in on an issue from several independent directions, you gain that much more expertise.

I've discussed secondary analysis in this chapter on survey research because it's the type of analysis most associated with the technique. However, there is no reason that the reanalysis of social research data needs to be limited to those collected in surveys. For example, when Dana Berkowitz and Maura Ryan (2011) set out to study how lesbian and gay parents deal with gender socialization for the adoptive children, they were able to find the qualitative data they needed in the qualitative interview records of two earlier studies of lesbian and gay parents. In taking a step beyond utilizing secondary studies, Nigel Fielding (2004) examined the possibilities for the archiving and reanalysis of qualitative data as well.

Ethics and Survey Research

Survey research almost always involves a request that people provide us with information about themselves that is not readily available. Sometimes, we ask for information (about attitudes and behaviors, for example) that would be embarrassing to the respondents if that information became publicly known. In some cases, such revelations could result in the loss of a job or a marriage. Hence, maintaining the norm of confidentiality, mentioned earlier in the book, is particularly important in survey research.

Another ethical concern relates to the possibility of psychological injury to respondents. Even if the information they provide is kept confidential, simply forcing them to think about some matters can be upsetting. Imagine asking people for their attitudes toward suicide when one of them has recently experienced the suicide of a family member or close friend. Or asking people to report on their attitudes about different racial groups, which may cause them to reflect on whether they might be racists or at least appear as such to the interviewers. The possibilities for harming survey respondents are endless. While this fact should not prevent you from doing surveys, it should increase your considered efforts to avoid the problem wherever possible.

MAIN POINTS

Introduction

- Survey research, a popular social research method, is the administration of questionnaires to a sample of respondents selected from some population.

Topics Appropriate for Survey Research

- Survey research is especially appropriate for making descriptive studies of large populations; survey data may be used for explanatory purposes as well.
- Questionnaires provide a method of collecting data by (1) asking people questions or (2) asking them to agree or disagree with statements representing different points of view.

Guidelines for Asking Questions

- Items in a questionnaire should follow several guidelines: (1) The form of the items should be appropriate to the project; (2) the items must be clear and precise; (3) the items should ask only about one thing (that is, double-barreled questions should be avoided); (4) respondents must be competent to answer the item; (5) respondents must be willing to answer the item; (6) questions should be relevant to the respondent; (7) items should ordinarily be short; (8) negative terms should be avoided so as not to confuse respondents; (9) the items should be worded to avoid biasing responses.
- Questions may be open-ended (respondents supply their own answers) or closed-ended (they select from a list of provided answers).

Questionnaire Construction

- The format of a questionnaire can influence the quality of data collected.

- A clear format for contingency questions is necessary to ensure that the respondents answer all the questions intended for them.
- The matrix question is an efficient format for presenting several items sharing the same response categories.
- The order of items in a questionnaire can influence the responses given.
- Clear instructions are important for getting appropriate responses in a questionnaire.
- Questionnaires should be pretested before being administered to the study sample.

Self-Administered Questionnaires

- Questionnaires are usually administered in one of three main ways: through self-administered questionnaires, face-to-face interviews, or telephone surveys. Researchers are exploring online surveys as well.
- It's generally advisable to plan follow-up mailings in the case of self-administered questionnaires, sending new questionnaires to those respondents who fail to respond to the initial appeal.
- Properly monitoring questionnaire returns will provide a good guide to when a follow-up mailing is appropriate.
- The ethics and efficacy of providing compensation has been a point of much debate.

Interview Surveys

- Interviewers must be neutral in appearance and actions; their presence in the data-collection process must have no effect on the responses given to questionnaire items.
- Interviewers must be carefully trained to be familiar with the questionnaire, to follow the question wording and question order exactly, and to record responses exactly as they are given.
- Interviewers can use probes to elicit an elaboration on an incomplete or ambiguous response. Probes should be neutral. Ideally, all interviewers should use the same probes.

Telephone Surveys

- Telephone surveys can be cheaper and more efficient than face-to-face interviews, and they can permit greater control over data collection.
- The development of computer-assisted telephone interviewing (CATI) is especially promising.
- Robo-polls are computer-executed phone surveys which involve no human interviewers

Online Surveys

- New technologies, including surveys over the Internet and those using mobile devices, offer

additional opportunities for social researchers. These methods, however, must be used with caution because respondents may not be representative of the intended population.

Mixed-Mode Surveys

- Sometimes it is appropriate to use more than one survey technique in a given study: telephone, mail, online.

Comparison of the Different Survey Methods

- The advantages of a self-administered questionnaire over an interview survey are economy, speed, lack of interviewer bias, and the possibility of anonymity and privacy to encourage candid responses on sensitive issues.

- The advantages of an interview survey over a self-administered questionnaire are fewer incomplete questionnaires and fewer misunderstood questions, generally higher completion rates, and greater flexibility in terms of sampling and special observations.

- The principal advantages of telephone surveys over face-to-face interviews are the savings in cost and time. There is also a safety factor: In-person interviewers might be required to conduct surveys in high-crime areas, which could pose a safety issue; telephone interviews, by design, eliminate such risks.

- Online surveys have many of the strengths and weaknesses of mail surveys. Although they're cheaper to conduct, ensuring that the respondents represent a more general population can be difficult.

Strengths and Weaknesses of Survey Research

- Survey research in general offers advantages in terms of economy, the amount of data that can be collected, and the chance to sample a large population. The standardization of the data collected represents another special strength of survey research.

- Survey research has several weaknesses: It is somewhat artificial, potentially superficial, and relatively inflexible. Using surveys to gain a full sense of social processes in their natural settings is difficult. In general, survey research is comparatively weak on validity and strong on reliability.

Secondary Analysis

- Secondary analysis provides social researchers with an important option for "collecting" data cheaply and easily but at a potential cost in validity.

Ethics and Survey Research

- Surveys often ask for private information, and researchers must keep such information confidential.

- Because asking questions can cause psychological discomfort or harm to respondents, the researcher should minimize this risk.

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.

bias	open-ended questions
closed-ended questions	probe
computer-assisted telephone interviewing (CATI)	questionnaire
	respondent
contingency question	response rate
interview	secondary analysis

PROPOSING SOCIAL RESEARCH: SURVEY RESEARCH

If you're planning a survey, you'll have already described the sampling you'll employ, and your discussion of measurement will have presented at least portions of your questionnaire. At this point you need to describe the type of survey you'll conduct: self-administered, telephone, face-to-face, or Internet. Whichever you plan, there will be numerous logistical details to spell out in the proposal. How will you deal with non-respondents, for example? Will you have follow-up mailing in a self-administered questionnaire, follow-up calls in a telephone survey, and so forth? Will you have a target completion rate?

In the case of interview surveys, you should say something about the way you'll select and train the interviewers. You should also say something about the time frame within which the survey will be conducted.

REVIEW QUESTIONS AND EXERCISES

1. For each of the following open-ended questions, construct a closed-ended question that could be used in a questionnaire.
   a. What was your family's total income last year?
   b. How do you feel about the space shuttle program?
   c. How important is religion in your life?

d. What was your main reason for attending college?

e. What do you feel is the biggest problem facing your community?

2. Construct a set of contingency questions for use in a self-administered questionnaire that would solicit the following information:

a. Is the respondent employed?

b. If unemployed, is the respondent looking for work?

c. If the unemployed respondent is not looking for work, is he or she retired, a student, or a homemaker?

d. If the respondent is looking for work, how long has he or she been looking?

3. Find a questionnaire printed in a magazine, newspaper, or on a website (for a reader survey, for example). Consider at least five of the questions in it and critique each one.

4. Look at your appearance right now. Identify aspects of your appearance that might create a problem if you were interviewing a general cross section of the public.

5. Locate a survey being conducted on the web. Briefly describe the survey and discuss its strengths and weaknesses.

Qualitative Field Research

CHAPTER OVERVIEW

Qualitative field research enables researchers to observe social life in its natural habitat: to go where the action is and watch. This type of research can produce a richer understanding of many social phenomena than can be achieved through other observational methods, provided that the researcher observes in a deliberate, well-planned, and active way.

Introduction

Several chapters ago, I suggested that you've been doing social research all your life. This idea should become even clearer as we turn to what probably seems like the most obvious method of making observations: qualitative field research. In a sense, we do field research whenever we observe or participate in social behavior and try to understand it, whether in a college classroom, in a doctor's waiting room, or on an airplane. Whenever we report our observations to others, we're reporting our field research efforts.

Such research is at once very old and very new in social science, stretching at least from the nineteenth-century studies of preliterate societies, through firsthand examinations of urban community life in the "Chicago School" of the 1930s and 1940s, to contemporary observations of chat-room interactions on the web. Many of the techniques discussed in this chapter have been used by social researchers for centuries. Within the social sciences, anthropologists are especially associated with this method and have contributed to its development as a scientific technique. Moreover, something similar to this method is employed by many people who might not, strictly speaking, be regarded as social science researchers. Newspaper reporters are one example; welfare department case workers are another.

Although these are "natural" activities, they are also skills to be learned and honed. This chapter discusses these skills in some detail, examining some of the major paradigms of field research and describing some of the specific techniques that make scientific field research more useful than the casual observation we all engage in.

I use the term *qualitative field research* to distinguish this type of observational method from methods designed to produce data appropriate for quantitative (statistical) analysis. Thus, surveys provide data from which to calculate the percentage unemployed in a population, mean incomes, and so forth. Field research more typically yields qualitative data: observations not easily reduced to numbers. Thus, for example,

a field researcher may note the "paternalistic demeanor" of leaders at a political rally or the "defensive evasions" of a public official at a public hearing without trying to express either the paternalism or the defensiveness as a numerical quantity or degree. Although field research can be used to collect quantitative data—for example, noting the number of interactions of various specified types within a field setting—typically, field research is qualitative.

Field observation also differs from some other models of observation in that it's not just a data-collecting activity. Frequently, perhaps typically, it's a theory-generating activity as well. As a field researcher, you'll seldom approach your task with precisely defined hypotheses to be tested. More typically, you'll attempt to make sense out of an ongoing process that cannot be predicted in advance—making initial observations, developing tentative general conclusions that suggest particular types of further observations, making those observations and thereby revising your conclusions, and so forth. In short, the alternation of induction and deduction discussed in Part 1 of this book is perhaps nowhere more evident and essential than in good field research. For expository purposes, however, this chapter focuses primarily on some of the theoretical foundations of field research and on techniques of data collection. Chapter 13 discusses how to analyze qualitative data.

Topics Appropriate for Field Research

One of the key strengths of field research is how comprehensive a perspective it can give researchers. By going directly to the social phenomenon under study and observing it as completely as possible, researchers can develop a deeper and fuller understanding of it. As such, this mode of observation is especially, though not exclusively, appropriate to research topics and social studies that appear to defy simple quantification. Field researchers may recognize several nuances of attitude and behavior that might escape researchers using other methods.

Field research is especially appropriate for the study of those attitudes and behaviors best understood within their natural setting, as opposed to the somewhat artificial settings of experiments and surveys. For example, field research provides a superior method for studying the dynamics of religious conversion at a revival meeting, just as a statistical analysis of membership rolls would be a better way of discovering whether men or women were more likely to convert.

Finally, field research is well suited to the study of social processes over time. Thus, the field researcher might be in a position to examine the rumblings and final explosion of a riot as events actually occur rather than afterward in a reconstruction of the events.

Or consider the insightful study of high school culture by Murray Milner Jr., appropriately entitled *Freaks, Geeks, and Cool Kids* (2004). Milner was interested in exploring two sets of questions: (1) why teen-agers behave in the ways they do and (2) how do their behaviors fit into the structure of the larger society?

Perhaps you can relate personally to one of the key starting points in Milner's study of teen-age life: the feeling that they are largely powerless in many aspects of their lives: "They must attend school for most of the day and they have only very limited influence on what happens there. They are pressured to learn complex and esoteric knowledge like algebra, chemistry, and European history, which rarely has immediate relevance to their day-to-day lives" (2004: 4).

Milner goes on to identify one area where teenagers do have, and exercise, a special kind of power:

> They do, however, have one crucial kind of power: the power to create an informal social world in which they evaluate one another. That is they can and do create their own status systems—usually based on criteria that are quite different from those promoted by parents or teachers.
>
> *(2004: 4)*

Status systems constitute a central concept for social scientists, and it was useful that Milner is also an expert on the Indian caste system, which figured into his examination and understanding of high school youth culture.

Other good places to apply field research methods include campus demonstrations,

courtroom proceedings, labor negotiations, public hearings, or similar events taking place within a relatively limited area and time. Several such observations must be combined in a more comprehensive examination over time and space.

In *Analyzing Social Settings* (2006: 123–132), John Lofland and his colleagues discuss several elements of social life appropriate to field research:

1. *Practices:* Various kinds of behavior, such as talking or reading a book
2. *Episodes:* A variety of events such as divorce, crime, and illness
3. *Encounters:* Two or more people meeting and interacting
4. *Roles and social types:* The analysis of the positions people occupy and the behavior associated with those positions: occupations, family roles, ethnic groups
5. *Social and personal relationships:* Behavior appropriate to pairs or sets of roles: mother–son relationships, friendships, and the like
6. *Groups and cliques:* Small groups, such as friendship cliques, athletic teams, and work groups
7. *Organizations:* Formal organizations, such as hospitals or schools
8. *Settlements and habitats:* Small-scale "societies" such as villages, ghettos, and neighborhoods, as opposed to large societies such as nations, which are difficult to study
9. *Social worlds:* Ambiguous social entities with vague boundaries and populations, such as "the sports world" and "Wall Street"
10. *Subcultures and lifestyles:* How large numbers of people adjust to life in groups such as a "ruling class" or an "urban underclass"

In all these social settings, field research can reveal things that would not otherwise be apparent. Here's a concrete example.

One issue I'm particularly interested in (Babbie 1985) is the nature of responsibility for public matters: Who's responsible for making the things that we share work? Who's responsible for keeping public spaces—parks, malls, buildings, and so on—clean? Who's responsible for seeing that broken street signs get fixed? Or, if a strong wind knocks over garbage cans and rolls them around the street, who's responsible for getting them out of the road?

On the surface, the answer to these questions is pretty clear. We have formal and informal agreements in our society that assign responsibility for these activities. Government custodians are responsible for keeping public places clean. Transportation department employees are responsible for the street signs, and perhaps the police are responsible for the garbage cans rolling around on a windy day. And when these responsibilities are not fulfilled, we tend to look for someone to blame.

What fascinates me is the extent to which the assignment of responsibility for public things to specific individuals not only relieves others of the responsibility but actually prohibits them from taking responsibility. It's my notion that it has become unacceptable for someone like you or me to take personal responsibility for public matters that haven't been assigned to us.

Let me illustrate what I mean. If you were walking through a public park and you threw down a bunch of trash, you'd discover that your action was unacceptable to those around you. People would glare at you, grumble to each other; perhaps someone would say something to you about it. Whatever the form, you'd be subjected to definite, negative sanctions for littering. Now here's the irony. If you were walking through that same park, came across a bunch of trash that someone else had dropped, and cleaned it up, it's likely that your action would also be unacceptable to those around you. You'd probably face negative sanctions for cleaning it up.

When I first began discussing this pattern with students, most felt the notion was absurd. Although we would be negatively sanctioned for littering, cleaning up a public place would obviously bring positive sanctions: People would be pleased with us for doing it. Certainly, all my students said they would be pleased if someone cleaned up a public place. It seemed likely that everyone else would be pleased, too, if we asked them how they would react to someone's cleaning up litter in a public place or otherwise taking personal responsibility for fixing some social problem.

To settle the issue, I suggested that my students start fixing the public problems they came across in the course of their everyday activities. As they did so, I asked them to note the answers to two questions:

1. How did they feel while they were fixing a public problem they had not been assigned responsibility for?

2. How did others around them react?

My students picked up litter, fixed street signs, put knocked-over traffic cones back in place, cleaned and decorated communal lounges in their dorms, trimmed trees that blocked visibility at intersections, repaired public playground equipment, cleaned public restrooms, and took care of a hundred other public problems that weren't "their responsibility."

Most reported feeling very uncomfortable doing whatever they did. They felt foolish, goody-goody, conspicuous, and all the other feelings that keep us from performing these activities routinely. In almost every case, their personal feelings of discomfort were increased by the reactions of those around them. One student was removing a damaged and long-unused newspaper box from the bus stop, where it had been a problem for months, when the police arrived, having been summoned by a neighbor. Another student decided to clean out a clogged storm drain on his street and found himself being yelled at by a neighbor who insisted that the mess should be left for the street cleaners. Everyone who picked up litter was sneered at, laughed at, and generally put down. One young man was picking up litter scattered around a trash can when a passerby sneered, "Clumsy!" It became clear to us that there are only three acceptable explanations for picking up litter in a public place:

1. You did it and got caught—somebody forced you to clean up your mess.

2. You did it and felt guilty.

3. You're stealing litter.

In the normal course of things, it's simply not acceptable for people to take responsibility for public things.

Clearly, we could not have discovered the nature and strength of agreements about taking personal responsibility for public things except through field research. Social norms suggest that taking responsibility is a good thing, sometimes referred to as good citizenship. Asking people what they thought about taking responsibility would have produced a solid consensus that it was good. Only going out into life, doing it, and

watching what happened gave us an accurate picture.

As an interesting footnote to this story, my students and I found that whenever people could get past their initial reactions and discover that the students were simply taking responsibility for fixing things for the sake of having them work, the passersby tended to assist. Although there are some very strong agreements making it "unsafe" to take responsibility for public things, the willingness of one person to rise above those agreements seemed to make it safe for others to do so, and they did.

Field research is not to be confused with journalism. Social scientists and journalists may use similar techniques, but they have quite a different relationship to data. For instance, individual interviewing is a common technique in journalism and sociology; nevertheless, sociologists are not simply concerned with reporting about a subject's attitude, belief, or experience. A sociologist's goal is to treat an interview as data that need to be analyzed to understand social life more generally.

Anne Byrne, John Canavan, and Michelle Millar (2009) suggest this distinction can go even deeper. The voice-centered relational (VCR) method focuses on who is speaking in communications and who is listening, taking accounts of the difference between the two actors and the impact of those differences. Often, the listener is the researcher. This approach shows up during interviews and during the analysis of transcripts. The authors say about their study that dealt with Irish teenagers,

> One of the challenging dimensions of the work was that it brought us face to face with a reality that demanded that we act with or on behalf of the teenagers. The work of relationship building is time consuming and energy sapping—many research approaches do not require the formation of "caring relationships" with the researched. Building relationships between old and young, from different class backgrounds and diverse life experiences require a sustained and shared commitment from all.
>
> *(2009: 75)*

Two important aspects of qualitative research need to be stressed. First, a wide range of studies fall under the umbrella "qualitative field research." As we'll see in this chapter, various epistemologies within different paradigms have quite different approaches to basic questions such as "What are data?" "How should we collect data?" and "How should we analyze data?" Second, we should remember that the questions we want to answer in our research determine the types of methods we need to use. A question such as "How do women construct their everyday lives in order to perform their roles as mothers, partners, and breadwinners?" could be addressed by in-depth interviews and direct observations. The assessment of advertising campaigns might profit from focus group discussions. In most cases, we'll find that researchers have alternate methods to choose from.

In summary, then, field research offers the advantage of probing social life in its natural habitat. Although some things can be studied adequately through questionnaires or in the laboratory, others cannot. And direct observation in the field lets researchers observe subtle communications and other events that might not be anticipated or measured otherwise.

Special Considerations in Qualitative Field Research

There are specific things to take into account in every research method, and qualitative field research is no exception. When you use field research methods, you're confronted with decisions about the role you'll play as an observer and your relations with the people you're observing. Let's examine some of the issues involved in these decisions.

The Various Roles of the Observer

In field research, observers can play any of several roles, including participating in what they want to observe (this was the situation of the students who fixed public things). In this chapter, I've used the term *field research* rather than the frequently used term *participant observation,* because field researchers need not always participate in what they're studying, though they usually do study it directly at the scene of the action.

As Catherine Marshall and Gretchen Rossman point out

> The researcher may plan a role that entails varying degrees of "participantness"—that is, the degree of actual participation in daily life. At one extreme is the full participant, who goes about ordinary life in a role or set of roles constructed in the setting. At the other extreme is the complete observer, who engages not at all in social interaction and may even shun involvement in the world being studied. And, of course, all possible complementary mixes along the continuum are available to the researcher.
>
> *(1995: 60)*

The complete participant, in this sense, may be a genuine participant in what he or she is studying (for example, a participant in a campus demonstration) or may pretend to be a genuine participant. In any event, whenever you act as the complete participant, you must let people see you only as a participant, not as a researcher. For instance, if you're using this technique to study a group made up of uneducated and inarticulate people, it would not be appropriate for you to talk and act like a university professor or student.

This type of research introduces an ethical issue, one on which social researchers themselves are divided. Is it ethical to deceive the people you're studying in the hope that they will confide in you as they will not confide in an identified researcher? Do the potential benefits to be gained from the research offset such considerations? Although many professional associations have addressed this issue, the norms to be followed remain somewhat ambiguous when applied to specific situations.

Related to this ethical consideration is a scientific one. No researcher deceives his or her subjects solely for the purpose of deception. Rather, it's done in the belief that the data will be more valid and reliable, that the subjects will be more natural and honest if they do not know the researcher is doing a research project. If the people being studied know they're being studied,

reactivity The problem that the subjects of social research may react to the fact of being studied, thus altering their behavior from what it would have been normally.

they might modify their behavior in a variety of ways. This is known as the problem of **reactivity**.

First, they might expel the researcher. Second, they might modify their speech and behavior to appear more "respectable" than would otherwise be the case. Third, the social process itself might be radically changed. Students making plans to burn down the university administration building, for example, might give up the plan altogether once they learn that one of their group is a social scientist conducting a research project.

On the other side of the coin, if you're a complete participant, you may affect what you're studying. Suppose, for example, that you're asked for your ideas about what the group should do next. No matter what you say, you will affect the process in some fashion. If the group follows your suggestion, your influence on the process is obvious. If the group decides not to follow your suggestion, the process whereby the suggestion is rejected may affect what happens next. Finally, if you indicate that you just don't know what should be done next, you may be adding to a general feeling of uncertainty and indecisiveness in the group.

Ultimately, anything the participant-observer does or does not do will have some effect on what's being observed; it's simply inevitable. More seriously, there is no complete protection against this effect, though sensitivity to the issue may provide a partial protection. (This influence, called the Hawthorne effect, was discussed more fully in Chapter 8.)

Because of these several considerations, ethical and scientific, the field researcher frequently chooses a different role from that of complete participant. You could participate fully with the group under study but make it clear that you were also undertaking research. As a member of the volleyball team, for example, you might use your position to launch a study in the sociology of sports, letting your teammates know what you're doing. There are dangers in this role also, however. The people being studied may shift much of their attention to the research project rather than focusing on the natural social process, making the process being observed no longer typical. Or, conversely, you yourself may come to identify too much with the interests and viewpoints of the participants. You may begin to "go native" and lose much of your scientific detachment.

At the other extreme, the complete observer studies a social process without becoming a part of it in any way. Quite possibly, because of the researcher's unobtrusiveness, the subjects of study might not realize they're being studied. Sitting at a bus stop to observe jaywalking at a nearby intersection is one example. Although the complete observer is less likely to affect what's being studied and less likely to "go native" than the complete participant, she or he is also less likely to develop a full appreciation of what's being studied. Observations may be more sketchy and transitory.

Fred Davis (1973) characterizes the extreme roles that observers might play as "the Martian" and "the Convert." The latter involves delving more and more deeply into the phenomenon under study, running the risk of "going native." We'll examine this risk further in the next section.

To appreciate the "Martian" approach, imagine that you were sent to observe some newfound life on Mars. Probably you would feel yourself inescapably separate from the Martians. Some social scientists adopt this degree of separation when observing cultures or social classes different from their own.

Marshall and Rossman (1995: 60–61) also note that the researcher can vary the amount of time spent in the setting being observed: You can be a full-time presence on the scene or just show up now and then. Moreover, you can focus your attention on a limited aspect of the social setting or seek to observe all of it—framing an appropriate role to match your aims.

When Jeffrey Kidder set out to study the culture of bike messengers in New York City, he found it appropriate to identify his research role to some of those he observed but not others:

> While I did have an academic motivation in working as a messenger, it should be made clear that my participation within the messenger world was neither forced nor faked. To the contrary, my lifelong interest in bicycles and alternative transportation melded seamlessly with the messenger lifestyle.
>
> During the course of my fieldwork, most of the messengers with whom I came in contact were unaware of my research; this was a matter of necessity. In New York City,

a messenger crosses paths with hundreds of messengers a day. The numerous individuals that helped form my understandings of messenger style could not all be approached to sign consent forms. Messengers with whom I had reoccurring contact were informed of my sociological interest.

> *(2005: 349)*

Different situations ultimately require different roles for the researcher. Unfortunately, there are no clear guidelines for making this choice—you must rely on your understanding of the situation and your own good judgment. In making your decision, however, you must be guided by both methodological and ethical considerations. Because these often conflict, your decision will frequently be difficult, and you may find sometimes that your role limits your study.

Relations to Subjects

Having introduced the different roles field researchers might play in connection with their observations, we now focus more specifically on how researchers may relate to the subjects of their study and to the subjects' points of view.

We've already noted the possibility of pretending to occupy social statuses we don't really occupy. Consider now how you would think and feel in such a situation.

Suppose you've decided to study a religious cult that has enrolled many people in your neighborhood. You might study the group by joining it or pretending to join it. Take a moment to ask yourself what the difference is between "really" joining and "pretending" to join. The main difference is whether or not you actually take on the beliefs, attitudes, and other points of view shared by the "real" members. If the cult members believe that Jesus will come next Thursday night to destroy the world and save the members of the cult, do you believe it or do you simply pretend to believe it?

Traditionally, social scientists have tended to emphasize the importance of "objectivity" in such matters. In this example, that injunction would be to avoid getting swept up in the beliefs of the group. Without denying the advantages associated with such objectivity, social scientists today also recognize the benefits gained by immersing themselves in the points of view

they're studying, what John Lofland and his colleagues (2006: 70) refer to as "selective competence" or "insider knowledge, skill, or understanding." Ultimately, you won't be able to fully understand the thoughts and actions of the cult members unless you can adopt their points of view as true—at least temporarily. To fully appreciate the phenomenon you've set out to study, you need to believe that Jesus is coming Thursday night. In some settings, this can also help you gain rapport with your subjects (see the discussion on rapport later in this chapter).

Adopting an alien point of view is an uncomfortable prospect for most people. It can be hard enough merely to learn about views that seem strange to you; you may sometimes find it hard just to tolerate certain views. But to take them on as your own can be ten times worse. Robert Bellah (1970, 1974) has offered the term *symbolic realism* to indicate the need for social researchers to treat the beliefs they study as worthy of respect rather than as objects of ridicule. The difficulty of adopting others' views led William Shaffir and Robert Stebbins to conclude that "fieldwork must certainly rank with the more disagreeable activities that humanity has fashioned for itself" (1991: 1).

There is, of course, a danger in adopting the points of view of the people you're studying. When you abandon your objectivity in favor of adopting such views, you lose the possibility of seeing and understanding the phenomenon within frames of reference unavailable to your subjects. On the one hand, accepting the belief that the world will end Thursday night allows you to appreciate aspects of that belief available only to believers; stepping outside that view, however, makes it possible for you to consider some reasons why people might adopt such a view. You may discover that some did so as a consequence of personal trauma (such as unemployment or divorce), whereas others were brought into the fold through their participation in particular social networks (for example, all their Facebook friends joined the cult). Notice that the cult members might disagree with those "objective" explanations, and you might not come up with them to the extent that you had operated legitimately within the group's views.

Anthropologists sometimes use the term *emic perspective* in reference to taking on the point of

view of those being studied. In contrast, the *etic perspective* maintains a distance from the native point of view in the interest of achieving more objectivity.

The apparent dilemma here is that both of these postures offer important advantages but also seem mutually exclusive. In fact, it's possible to assume both postures. Sometimes you can simply shift viewpoints at will. When appropriate, you can fully assume the beliefs of the cult; later, you can step outside those beliefs (more accurately, you can step inside the viewpoints associated with social science). As you become more adept at this kind of research, you may come to hold contradictory viewpoints simultaneously, rather than switching back and forth.

During my study of trance channeling—in which people allow spirits to occupy their bodies and speak through them—I found I could participate fully in channeling sessions without becoming alienated from conventional social science. Rather than "believing" in the reality of channeling, I found it possible to suspend beliefs in that realm: neither believing it to be genuine (like most of the other participants) nor disbelieving it (like most scientists). Put differently, I was open to either possibility. Notice how this differs from our normal need to "know" whether such things are legitimate or not.

Social researchers often refer to the concerns just discussed as a matter of *reflexivity,* in the sense of things acting on themselves. Thus, your own characteristics can affect what you see and how you interpret it. The issue is broader than that, however, and applies to the subjects as well as to the researcher. Imagine yourself interviewing a homeless person (1) on the street, (2) in a homeless shelter, or (3) in a social welfare office. The research setting could affect the person's responses. In other words, you might get different results depending on where you conducted the interview. Moreover, you might act differently as a researcher in those different settings. If you reflect on this issue, you'll be able to identify other aspects of the research encounter that complicate the task of "simply observing what's so."

The problem we've just been discussing could be seen as psychological, occurring mostly inside the researchers' or subjects' heads. There is a corresponding problem at a social level, however.

When you become deeply involved in the lives of the people you're studying, you're likely to be moved by their personal problems and crises. Imagine, for example, that one of the cult members becomes ill and needs a ride to the hospital. Should you provide transportation? Sure. Suppose someone wants to borrow money to buy a stereo. Should you loan it? Probably not. Suppose they need the money for food?

There are no black-and-white rules for resolving situations such as these, but you should realize that you'll need to deal with them regardless of whether or not you reveal that you're a researcher. Such problems do not tend to arise in other types of research—surveys and experiments, for example—but they are part and parcel of field research.

Caroline Knowles (2006) raises a somewhat different issue with regard to the researcher's relationship to subjects in the field. In her interview study of British expatriates living in Hong Kong, she noticed that some were particularly difficult for her to deal with. When she found herself writing research notes explaining why the project would not profit from her interviewing them further, she forced herself to look more deeply into the interactional dynamics in question—with an emphasis on her side of the relationships. She examined *why* certain informants made her uncomfortable and then pressed through the discomfort to continue interviewing. She found that factors such as the attitudes they expressed, their rude interaction styles, and the nature of the relationship she was establishing with them contributed to her reaction. In the end, she gained a much deeper understanding of her subjects than would have been possible if she had limited herself to those who were cooperative and nice.

Similarly, Alex Broom, Kelly Hand, and Philip Tovey (2009) examined the impact of gender when conducting in-depth interviews with cancer patients. Did it matter whether patients were interviewed by someone of the same or of the opposite gender? It did. Prostate cancer patients were more graphic in describing their experiences to a male interviewer than to a woman. Similarly, a breast-cancer patient's feelings of disfigurement, for example, were expressed differently to male and female interviewers. Before you decide that gender-matching is the best policy, notice that a cancer patient's overall experience includes same-gender and opposite-gender relations. As I have said frequently in this book, the impact of the observer, whether in experiments, surveys, or field research often cannot be avoided, but we can be conscious of it and take it into account in understanding what we have observed.

This discussion of the field researcher's relationships to subjects flies in the face of the usual view of "scientific objectivity." Before concluding this section, let's take the issue one step further.

In the conventional view of science, differences of power and status separate the researcher from the subjects of research. When we discussed experimental designs in Chapter 8, for example, it was obvious who was in charge: the experimenter, who organized things and told the subjects what to do. Often the experimenter was the only person who even knew what the research was really about. Something similar might be said about survey research. The person running the survey designs the questions, decides who will be selected for questioning, and analyzes the data collected.

Sociologists often look at these sorts of relationships as power or status relationships. In experimental and survey designs, the researcher clearly has more power and a higher status than the people being studied do. The researchers have a special knowledge that the subjects don't enjoy. They're not so crude as to say they're superior to their subjects, but there is a sense in which that's implicitly assumed. (Notice that there is a similar, implicit assumption about the writers and readers of textbooks.)

In field research, such assumptions can be problematic. When the early European anthropologists set out to study what were originally called "primitive" societies, there was no doubt that the anthropologists knew best. Whereas the natives "believed" in witchcraft, for example, the anthropologists "knew" it wasn't really true. Whereas the natives said some of their rituals would appease the gods, the anthropologists explained that the "real" functions of these rituals were the creation of social identity, the establishment of group solidarity, and so on.

Giampietro Gobo (2011) sensitizes us to the cultural roots (and limits) of commonly used social research techniques. For the most part,

Research in Real Life

Class Acts: Service and Inequality in Luxury Hotels

What could seem like a clearer status relationship than between a guest in a luxury hotel and the room service and other staff who serve that guest's needs? In fact, Rachel Sherman has found a far more complex process than you might imagine. She is particularly interested in how service workers balance their relationships with management and their relationships with guests. Unlike manufacturing workers, the hotel service staff must deal with both supervisors and consumers, even when the demands of the two conflict. In part, she discovered that service workers in hotels often receive more discretion regarding how to serve guests' needs than we might expect. This has a positive impact on the worker's sense of self as well as providing a good experience for guests.

Sherman's observations and conclusions came from months spent as a service worker in two luxury hotels. She made her research identity known to management and was able to move around through many of the different service jobs: making reservations, delivering room-service meals, parking cars, carrying bags, housekeeping, and many other tasks that the guests in luxury hotels expect. Her immersion in the research allowed her access to data she would not have found otherwise.
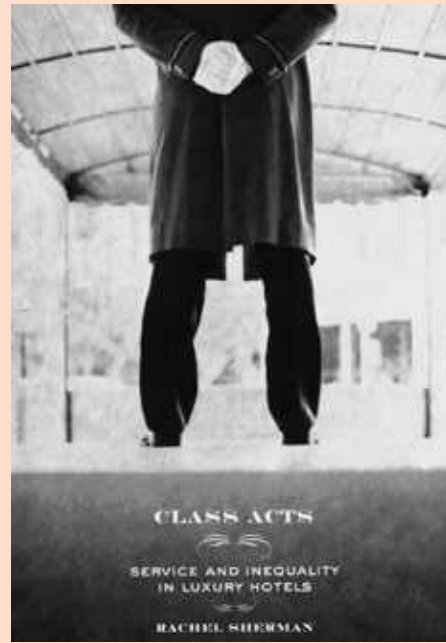
Walter B. McKenzie/The Image Bank/Getty Images

---

those roots are embedded in American and European cultures. By contrast, he notes that

> If one wanted to undertake an ethnomethodological study in the Maya villages of Zincatela and San Juan Chamula, in Chiapas (Mexico), one should be careful about the following: no photos are allowed in the church or of people in this area due to the belief that photos will capture the soul, and if someone leaves the village with the photo then their soul goes with them.
>
> *(2011: 417)*

Gobo also points out some of the implicit assumptions that lie behind the use of standardized or unstructured interviews, such as "the sense that it is acceptable to have conversations with strangers (interviewers)"; "the ability on the part of the interviewee to speak for himself, and an awareness of himself as an autonomous and independent individual"; and "experience in giving information in telephone interviews without seeing the face of the interviewer" (2011: 424). These are foreign concepts for many who were raised outside of Europe and America.

The more social researchers have gone into the field to study their fellow humans face-to-face, however, the more they have become conscious of these implicit assumptions about researcher superiority, and the more they have considered alternatives. As we turn now to the various paradigms of field research, we'll see some of the ways in which that ongoing concern has worked itself out. See the Research in Real Life box, "Class Acts: Service and Inequality in Luxury Hotels," for an example of field research on status.

Some Qualitative Field Research Paradigms

Although I've described field research as simply going where the action is and observing it, there are actually many different approaches to this research method. This section examines several field research paradigms: naturalism, ethnomethodology, grounded theory, case studies and the extended case method, institutional ethnography, and participatory action research. Although

this survey won't exhaust the variations on the method, it should give you a broad appreciation of the possibilities.

It's important to recognize that there is no single method appropriate to a particular paradigm. You could do ethnomethodology or institutional ethnography by analyzing court hearings or conducting group interviews, for example. The important distinctions of this section are epistemological, having to do with what data mean, regardless of how they were collected.

Naturalism

Naturalism is an old tradition in qualitative research. The earliest field researchers operated on the positivist assumption that social reality was "out there," ready to be naturally observed and reported by the researcher as it "really is" (Gubrium and Holstein 1997). This tradition started in the 1930s and 1940s at the University of Chicago's sociology department, whose faculty and students fanned out across the city to observe and understand local neighborhoods and communities. The researchers of that era and their research approach are now often referred to as the Chicago School.

One of the earliest and best-known studies that illustrates this research tradition is William Foote Whyte's ethnography of Cornerville, an Italian American neighborhood, in his book *Street Corner Society* (1943). An **ethnography** is a study that focuses on detailed and accurate description rather than explanation. Like other naturalists, Whyte believed that in order to learn fully about social life on the streets, he needed to become more of an insider. He made contact with "Doc," his key informant, who appeared to be one of the street-gang leaders. Doc let Whyte enter his world, and Whyte got to participate in the activities of the people of Cornerville. His study offered something that surveys could not: a richly detailed picture of life among the Italian immigrants of Cornerville.

An important feature of Whyte's study is that he reported the reality of the people of Cornerville on their terms. The naturalist approach is based on telling "their" stories the way they "really are," not the way the ethnographer understands "them." The narratives collected by Whyte are taken at face value as the social "truth" of the Cornerville residents.

About 40 years later, David Snow and Leon Anderson (1987) conducted exploratory field research into the lives of homeless people in Austin, Texas. Their main task was to understand how the homeless construct and negotiate their identity while knowing that the society they live in attaches a stigma to homelessness. Snow and Anderson believed that, to achieve this goal, the collection of data had to arise naturally. Like Whyte in *Street Corner Society*, they found some key informants whom they followed in their everyday journeys, such as at their day-labor pickup sites or under bridges. Snow and Anderson chose to memorize the conversations they participated in or the "talks" that homeless people had with each other. At the end of the day, the two researchers debriefed and wrote detailed field notes about all the "talks" they encountered. They also recorded in-depth interviews with their key informants.

Snow and Anderson reported "hanging out" with homeless people over the course of 12 months for a total of 405 hours in 24 different settings. Out of these rich data, they identified three related patterns in homeless people's conversations. First, the homeless showed an attempt to "distance" themselves from other homeless people, from the low-status job they currently had, or from the Salvation Army they depended on. Second, they "embraced" their street-life identity—their group membership or a certain belief about why they are homeless. Third, they told "fictive stories" that always contrasted with their everyday life. For example, they would often say that they were making much more money than they really were, or even that they were "going to be rich."

In a more recent attempt to understand the daily lives of people cordoned off from the American Dream, Alice Goffman (2014) spent six years living in a distressed neighborhood in Philadelphia, sharing an apartment with young people living near or beyond the edge of the law.

naturalism An approach to field research based on the assumption that an objective social reality exists and can be observed and reported accurately.

ethnography A report on social life that focuses on detailed and accurate description rather than explanation.

In particular, Goffman saw firsthand how nearly everyone has lost in the government's "war on drugs," instigated by President Nixon in 1971. No one would claim this campaign has resulted in the disappearance of illegal drugs from American society, and the unintended consequences have been enormous. Young minorities in poor, urban neighborhoods are common casualties in this war as is the case in more-conventional wars. Participant observation offered a way to see how the war on drugs was being fought on the ground.

Goffman learned what it is like to live in an environment largely organized around the continuing system of antidrug enforcement. The young men in her apartment were constantly harassed by the police, often arrested, tried, and jailed. As the title of her book suggests, they tended to live life "on the run."

Even those, such as Goffman herself, who were not dealing drugs were nonetheless subject to police attention. She found herself pulled in for questioning and pressured to testify about the illegal activities of the young men she knew in the neighborhood. As she learned, this was a common experience for the family members of suspected drug dealers. For all residents, criminal or not, neighborhood life was organized around police surveillance and enforcement. Some members of the neighborhood engaged in criminal activities, some were frequent victims of crime, and many fit in both categories. Goffman was struck by the virtual impossibility of ever escaping from such a depressing life situation.

While ethnographers seek to discover and understand the patterns of living among those they are studying, Mitchell Duneier (1999) has warned against what he calls the "ethnographic fallacy." This refers to an overgeneralization and oversimplification of the patterns observed. Despite the existence of patterns within groups, there is also diversity, and you need to be wary of broad assertions suggesting that "the poor," "the French," or "cheerleaders" act or think

in certain ways as though all members of the group do so.

Whereas this chapter aims at introducing some of the different approaches available to you in qualitative field research, please realize that this discussion of ethnography merely sketches some of the many avenues social researchers have established. If you're interested in this general approach, you might want to explore the idea of *virtual ethnography,* which uses ethnographic techniques for inquiry into online social networks or communities. Or, in a different direction, *autoethnography* intentionally assumes a personal stance, breaking with the general proscription against the researcher getting involved at that level. Lest autoethnography seem a simple and/or trivial undertaking, you might look at Sarah Wall's 2008 article, "Easier Said than Done: Writing an Autoethnography."

You can learn more about these variants on ethnography by searching the web or your campus library. A later section of this chapter will examine *institutional ethnography,* which links individuals and organizations.

In Chapter 9, we saw how the Internet is affecting survey research. Eric Anderson (2005) used the Internet to launch a qualitative, in-depth interviewing study of male cheerleaders. He began by using a search engine to identify men whose online profiles contained an interest in cheerleading. He contacted them via instant messaging and requested recorded telephone interviews.

Anderson then used snowball sampling to increase the number of cheerleaders to study. This is just another example of the wide variety of venues for ethnographic study.

Ethnomethodology

Ethnomethodology, which I introduced as a research paradigm in Chapter 2, is a unique approach to qualitative field research. It has its roots in the philosophical tradition of phenomenology, which can explain why ethnomethodologists are skeptical about the way people report their experience of reality (Gubrium and Holstein 1997). Alfred Schutz (1967, 1970), who introduced phenomenology, argued that reality was socially constructed rather than being "out there" for us to observe. People describe their world not "as it is" but "as they make sense of it." Thus, phenomenologists would argue that

ethnomethodology An approach to the study of social life that focuses on the discovery of implicit, usually unspoken assumptions and agreements; this method often involves the intentional breaking of agreements as a way of revealing their existence.

Whyte's street-corner men were describing their gang life as it made sense to them. Their reports, however, would not tell us how and why it made sense to them. For this reason, researchers cannot rely on their subjects' stories to depict social realities accurately.

Whereas traditional ethnographers believe in immersing themselves in a particular culture and reporting their informants' stories as if they represented reality, phenomenologists see a need to "make sense" out of the informants' perceptions of the world. Following in this tradition, some field researchers have felt the need to devise techniques that reveal how people make sense of their everyday world. As we saw in Chapter 2, the sociologist Harold Garfinkel suggested that researchers *break the rules* so that people's taken-for-granted expectations would become apparent. This is the technique that Garfinkel called ethnomethodology.

Garfinkel became known for engaging his students to perform a series of what he called "breaching experiments" designed to break away from the ordinary (Heritage 1984). For instance, Garfinkel (1967) asked his students to do a "conversation clarification experiment." Students were told to engage in an ordinary conversation with an acquaintance or a friend and to ask for clarification about any of this person's statements. Through this technique, they uncovered elements of conversation that are normally taken for granted. Here are two examples of what Garfinkel's students reported (1967: 42):

> Case 1
> The subject [S] was telling the experimenter [E], a member of the subject's car pool, about having had a flat tire while going to work the previous day.
> [S] I had a flat tire.
> [E] What do you mean, you had a flat tire?
> She appeared momentarily stunned. Then she answered in a hostile way: [S] "What do you mean, 'What do you mean?' A flat tire is a flat tire. That is what I meant. Nothing special. What a crazy question."

> Case 6
> [S] How are you?
> [E] How I am in regard of what? My health, my finances, my school work, my peace of mind, my . . . ?

> [S] (Red in the face and suddenly out of control.) Look I was just trying to be polite. Frankly, I don't give a damn how you are.

By setting aside or "bracketing" their expectations from these everyday conversations, the experimenters made visible the subtleties of mundane interactions. For example, although "How are you?" has many possible meanings, none of us have any trouble knowing what it means in casual interactions, as the unsuspecting subject revealed in his final comment.

Ethnomethodologists, then, are not simply interested in subjects' perceptions of the world. In these cases, we could imagine that the subjects may have thought that the experimenters were rude, stupid, or arrogant. The conversation itself, not the informants, is the object of ethnomethodological studies. In general, ethnomethodology focuses on the "underlying patterns" of interactions that regulate our everyday lives.

Ethnomethodologists believe that researchers who use a naturalistic analysis "[lose] the ability to analyze the commonsense world and its culture if [they use] analytical tools and insights that are themselves part of the world or culture being studied" (Gubrium and Holstein 1997: 43). D. L. Wieder provides an excellent example of how different a naturalistic approach is from an ethnomethodological approach (Gubrium and Holstein 1997). In his study *Language and Social Reality: The Case of Telling the Convict Code* (1988), Wieder started to approach convicts in a halfway house in a traditional ethnographic style: He was going to become an insider by befriending the inmates and by conducting participant observations. He took careful notes and recorded interactions among inmates and between inmates and staff. His first concern was to describe the life of the convicts of the halfway house the way it "really was" for them. Wieder's observations allowed him to report on a "convict code" that he thought was the source of the deviant behavior expressed by the inmates toward the staff. This code, which consisted of a series of rules such as "Don't kiss ass," "Don't snitch," and "Don't trust the staff," was followed by the inmates who interfered with the staff members' attempts to help them make the transition between prison and the community.

It became obvious to Wieder that the code was more than an explanation for the convicts'
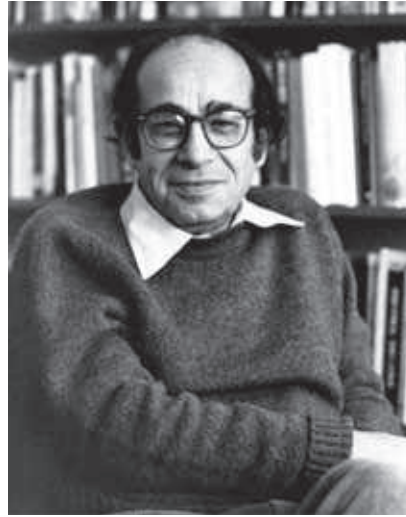
deviant behavior; it was a "method of moral persuasion and justification" (Wieder 1988: 175). At this point he changed his naturalistic approach to an ethnomethodological one. Whereas naturalistic field researchers aim to understand social life as the participants understand it, ethnomethodologists are more intent on identifying the methods through which understanding occurs. In the case of the convict code, Wieder came to see that convicts used the code to make sense of their own interactions with other convicts and with the staff. The ethnography of the halfway house thus shifted to an ethnography of the code. For instance, the convicts would say, "You know I won't snitch," referring to the code as a way to justify their refusal to answer Wieder's question (168). According to Wieder, the code "operated as a device for stopping or changing the topic of conversation" (175). Even the staff would refer to the code to justify their reluctance to help the convicts. Although the code was something that constrained behavior, it also functioned as a tool for the control of interactions.

Grounded Theory

Grounded theory originated from the collaboration of Barney Glaser and Anselm Strauss, sociologists who brought together two main traditions of research: positivism and interactionism. Essentially, **grounded theory** is the attempt to derive theories from an analysis of the patterns, themes, and common categories discovered in observational data. The first major presentation of this method can be found in Glaser and Strauss's book, *The Discovery of Grounded Theory* (1967). Grounded theory can be described as an approach that attempts to combine a naturalist approach with a positivist concern for a "systematic set of procedures" in doing qualitative research.

In *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory* (1998: 43–46), Anselm Strauss and Juliet Corbin have suggested that grounded theory allows

© Lynda Koolish

Anselm L. Strauss, a pioneer qualitative researcher, was a principal founder of the Grounded Theory Method.

the researcher to be scientific and creative at the same time, as long as the researcher follows these guidelines:

- *Think comparatively:* The authors suggest that it is essential to compare numerous incidents as a way of avoiding the biases that can arise from interpretations of initial observations.

- *Obtain multiple viewpoints:* In part this refers to the different points of view of participants in the events under study, but Strauss and Corbin suggest that different observational techniques may also provide a variety of viewpoints.

- *Periodically step back:* As data accumulate, you'll begin to frame interpretations about what is going on, and it's important to keep checking your data against those interpretations. As Strauss and Corbin say, "The data themselves do not lie" (1998: 45).

- *Maintain an attitude of skepticism:* As you begin to interpret the data, you should regard all those interpretations as provisional, using new observations to test those interpretations, not just confirm them.

- *Follow the research procedures:* Grounded theory allows for flexibility in data collection as theories evolve, but Strauss and Corbin stress that three techniques are essential: "making comparisons, asking questions, and sampling" (1998: 46).

grounded theory An inductive approach to the study of social life that attempts to generate a theory from the constant comparing of unfolding observations. This is very different from hypothesis testing, in which theory is used to generate hypotheses to be tested through observations.

Although a review of the existing research literature is an early step in most research methods—toward the goal of learning what has been learned so far—the initial development of grounded theory argued specifically *against* this practice. Glaser and Strauss feared that grounding yourself in what has already been learned would create expectations that would constrain what your research would look at, what you would see, and how you would interpret the data. Rather, they urged that categories and patterns be allowed to emerge from the new data. Once the data collection was ended, a review of the literature would provide another opportunity for comparison.

As Ciarán Dunne (2011) has detailed, this position regarding literature review is one now debated among grounded theorists. Strauss, himself, modified his own opinion on the matter, acknowledging it might be appropriate to do an early literature review in some cases. Glaser, on the other hand, has maintained his original position. In any event, the initial concerns about preconceptions possibly blinding researchers to new discoveries can apply to any research method you might choose.

Grounded theory emphasizes research procedures. In particular, systematic coding is important for achieving validity and reliability in the data analysis. Because of this somewhat positivistic view of data, grounded theorists are quite open to the use of qualitative studies in conjunction with quantitative ones. Here are two examples of the implementation of this approach.

Studying Academic Change

Clifton Conrad's (1978) study of academic change in universities is an early example of the grounded theory approach. Conrad hoped to uncover the major sources of changes in academic curricula and at the same time understand the process of change. Using the grounded theory idea of theoretical sampling—whereby groups or institutions are selected on the basis of their theoretical relevance—Conrad chose four universities for the purpose of his study. In two, the main vehicle of change was the formal curriculum committee; in the other two, the vehicle of change was an ad hoc group.

Conrad explained, step by step, the advantage of using the grounded theory approach in building his theory of academic change. He described the process of systematically coding data in order to create categories that must "emerge" from the data and then assessing the fitness of these categories with each other. Going continuously from data to theory and theory to data allowed him to reassess the validity of his initial conclusions about academic change.

For instance, it first seemed that academic change was mainly caused by an administrator who was pushing for it. By reexamining the data and looking for more-plausible explanations, Conrad found the pressure of interest groups a more convincing source of change. The emergence of these interest groups actually allowed the administrator to become an agent of change.

Assessing how data from each of the two types of universities fit with the other helped refine the theory building. Conrad concluded that changes in university curricula are based on the following process: Conflict and interest groups emerge because of internal and external social structural forces; they push for administrative intervention and recommendation to make changes in the current academic program; these changes are then made by the most powerful decision-making body.

Shopping Romania

Much has been written about large-scale changes caused by the shift from socialism to capitalism in the former USSR and its Eastern European allies. Patrick Jobes and his colleagues (1997) wanted to learn about the transition on a smaller scale among average Romanians. They focused on the task of shopping.

Noting that shopping is normally thought of as a routine, relatively rational activity, the researchers suggested that it could become a social problem in a radically changing economy. They used the Grounded Theory Method to examine Romanian shopping as a social problem, looking for the ways in which ordinary people solved the problem.

Their first task was to learn something about how Romanians perceived and understood the task of shopping. The researchers—participants in a social problems class—began by interviewing 40 shoppers and asking whether they had experienced problems in connection with shopping and what actions they had taken to cope with those problems.

Once the initial interviews were completed, the researchers reviewed their data, looking for categories of responses—the shoppers' most common problems and solutions. One of the most common problems was a lack of money. This led to the researchers' first working hypothesis: The "socio-economic position of shoppers would be associated with how they perceived problems and sought solutions" (1997: 133). This and other hypotheses helped the researchers focus their attention on more-specific variables in subsequent interviewing.

As they continued, they also sought to interview other types of shoppers. When they interviewed students, for example, they discovered that different types of shoppers were concerned with different kinds of goods, which in turn affected the problems faced and the solutions tried.

As the researchers developed additional hypotheses in response to the continued interviewing, they also began to develop a more or less standardized set of questions to ask shoppers. Initially, all the questions were open-ended, but they eventually developed closed-ended items as well.

This study illustrates the key, inductive principles of grounded theory: Data are collected in the absence of hypotheses. The initial data are used to determine the key variables as perceived by those being studied, and hypotheses about relationships among the variables are similarly derived from the data collected. Continuing data collection yields refined understanding and, in turn, sharpens the focus of data collection itself.

Case Studies and the Extended Case Method

Social researchers often speak of case studies. A **case study** focuses attention on a single instance of some social phenomenon, such as a village, a family, or a juvenile gang. As Charles Ragin and Howard Becker (1992) point out, there is little consensus on what may constitute a "case," and

case study The in-depth examination of a single instance of some social phenomenon, such as a village, a family, or a juvenile gang.

extended case method A technique developed by Michael Burawoy in which case study observations are used to discover flaws in and to improve existing social theories.

the term is used broadly. The case being studied, for example, might be a period of time rather than a particular group of people. The limitation of attention to a particular instance of something is the essential characteristic of the case study.

The chief purpose of case studies may be descriptive, as when an anthropologist describes the culture of a preliterate tribe. Or the in-depth study of a particular case can yield explanatory insights, as when the community researchers Robert and Helen Lynd (1929, 1937) and W. Lloyd Warner (1949) sought to understand the structure and process of social stratification in small-town USA.

Case study researchers may seek only an idiographic understanding of the particular case under examination, or—as we've seen with grounded theory—case studies can form the basis for the development of more-general, nomothetic theories.

Michael Burawoy and his colleagues (1991) have suggested a somewhat different relationship between case studies and theory. For them, the **extended case method** has the purpose of discovering flaws in, and then modifying, existing social theories. This approach differs importantly from some of the others already discussed.

Whereas traditional grounded theorists seek to enter the field with no preconceptions about what they'll find, Burawoy suggests just the opposite: to try "to lay out as coherently as possible what we expect to find in our site *before* entry" (Burawoy et al. 1991: 9). Burawoy sees the extended case method as a way to rebuild or improve theory instead of approving or rejecting it. Thus, he looks for all the ways in which observations conflict with existing theories and what he calls "theoretical gaps and silences" (1991: 10). This orientation to field research implies that knowing the literature beforehand is actually a must for Burawoy and his colleagues, whereas some grounded theorists would worry that knowing what others have concluded might bias their observations and theories.

To illustrate the extended case method, I'll present two examples of studies by Burawoy's students.

Teacher–Student Negotiations

Leslie Hurst (1991) set out to study the patterns of interaction between teachers and students of a

junior high school. She went into the field armed with existing contradictory theories about the "official" functions of the school. Some theories suggested that the purpose of schools is to promote social mobility, whereas others suggested that schools mainly reproduce the status quo in the form of a stratified division of labor. The official roles assigned to teachers and students could be interpreted in terms of either view.

Hurst was struck, however, by the contrast between these theories and the types of interactions she observed in the classroom. In her own experiences as a student, teachers had total rights over the minds, bodies, and souls of their pupils. She observed something quite different at a school in a lower-middle-class neighborhood in Berkeley, California, where she volunteered as a tutor. She had access to the classroom of Mr. Henry (an eighth-grade English teacher) as well as other teachers' classrooms, the lunchroom, and English department meetings. She wrote field notes based on the negotiations between students and teachers. She explained the nature of the student–teacher negotiations she witnessed by focusing on the separation of functions among the school, the teacher, and the family.

In Hurst's observation, the school fulfilled the function of controlling its students' "bodies"—for example, by regulating their general movements and activities within the school. The students' "minds" were to be shaped by the teacher, whereas students' families were held responsible for their "souls"; that is, families were expected to socialize students regarding personal values, attitudes, sense of property, and sense of decorum. When students don't come to school with these values in hand, the teacher, according to Hurst, "must first negotiate with the students some compromise on how the students will conduct themselves and on what will be considered classroom decorum" (1991: 185).

Hurst explained that the constant bargaining between teachers and students is an expression of the separation between "the body," which is the school's concern, and "the soul" as family domain. The teachers, who had limited sanctioning power to control their students' minds in the classroom, were using forms of negotiations with students so that they could "control . . . the student's body and sense of property" (1991: 185), or as Hurst defines it, "babysit" the student's body and soul.

Hurst says her method differs from the traditional sociological perspectives as follows:

> I do not approach schools with a futuristic eye. I do not see the school in terms of training, socializing, or slotting people into future hierarchies. To approach schools in this manner is to miss the negotiated, chaotic aspects of the classroom and educational experience. A futurist perspective tends to impose an order and purpose on the school experience, missing its day-to-day reality.
>
> *(1991: 186)*

In summary, what emerges from Hurst's study is an attempt to improve the traditional sociological understanding of education by adding the idea that classroom, school, and family have separate functions, which in turn can explain the emergence of "negotiated order" in the classroom.

The Fight against AIDS

Katherine Fox (1991) set out to study an agency whose goal was to fight the AIDS epidemic by bringing condoms and bleach for cleaning needles to intravenous drug users. It's a good example of finding the limitations of well-used models of theoretical explanation in the realm of understanding deviance—specifically, the "treatment model" that predicted that drug users would come to the clinic and ask for treatment. Fox's interactions with outreach workers—most of whom were part of the community of drug addicts or former prostitutes—contradicted that model.

To begin, it was necessary to understand the drug users' subculture and use that knowledge to devise more-realistic policies and programs. The target users had to be convinced, for example, that the program workers could be trusted, that they were really interested only in providing bleach and condoms. The target users needed to be sure they were not going to be arrested.

Fox's field research didn't stop with an examination of the drug users. She also studied the agency workers, discovering that the outreach program meant different things to the research directors and the outreach workers. Some of the volunteers who were actually providing the bleach and condoms were frustrated about the minor changes they felt they could make. Many

thought the program was just a bandage on the AIDS and drug-abuse problems. Some resented having to take field notes. Directors, on the other hand, needed reports and field notes so that they could validate their research in the eyes of the federal and state agencies that financed the project. Fox's study showed how the AIDS research project developed the bureaucratic inertia typical of established organizations: Its goal became that of sustaining itself.

Both of these studies illustrate how the extended case method can operate. The researcher enters the field with full knowledge of existing theories but aims to uncover contradictions that require the modification of those theories.

In all this, however, it is important to consider the issue of generalizability, especially when causal conclusions are reached. This is especially problematic when journalists engage in "case studies." With the writer having reached a conclusion about a particular cause-effect relationship, a discerning social scientist will question whether the journalist has sought out other cases that confirm the identified relationship and simply ignored those that contradict it. Properly analyzed case studies are careful to guard against that risk—and this pitfall can be reduced when more than one case is studied in depth, a method called the *comparative* case study. You can find examples of this in the discussion of comparative and historical methods in Chapter 11 of this book.

Institutional Ethnography

Institutional ethnography is an approach originally developed by Dorothy Smith (1978) to better understand women's everyday experiences by discovering the power relations that shape those experiences. Today this methodology has been extended to the ideologies that shape the experiences of any oppressed subjects.

Smith and other sociologists believe that if researchers ask women or other members of subordinated groups about "how things work," they

institutional ethnography A research technique in which the personal experiences of individuals are used to reveal power relationships and other characteristics of the institutions within which they operate.

Courtesy of Dr. Dorothy E. Smith

Dorothy Smith, a pioneering social researcher and founder of institutional ethnography.

can discover the institutional practices that shape their realities (M. L. Campbell 1998; D. Smith 1978). The goal of such inquiry is to uncover forms of oppression that more-traditional types of research often overlook.

Dorothy Smith's methodology is similar to ethnomethodology in the sense that the subjects themselves are not the focus of the inquiry. The institutional ethnographer starts with the personal experiences of individuals but proceeds to uncover the institutional power relations that structure and govern those experiences. In this process, the researcher can reveal aspects of society that would have been missed by an inquiry that began with the official purposes of institutions.

This approach links the "microlevel" of everyday personal experiences with the "macrolevel" of institutions. As M. L. Campbell puts it,

> Institutional ethnography, like other forms of ethnography, relies on interviewing, observations and documents as data. Institutional ethnography departs from other ethnographic approaches by treating those data not as the topic or object of interest, but as "entry" into the social relations of the setting. The idea is to tap into people's expertise.
>
> *(1998: 57)*

Here are two examples of this approach.

Mothering, Schooling, and Child Development

Our first example of institutional ethnography is a study by Alison Griffith (1995), who collected data with Dorothy Smith on the relationship among mothering, schooling, and children's development. Griffith started by interviewing mothers from three cities of southern Ontario about their everyday work of creating a relationship between their families and the school. This was the starting point for other interviews with parents, teachers, school administrators, social workers, school psychologists, and central office administrators.

In her findings, Griffith explained how the discourse about mothering had shifted its focus over time from a mother–child interaction to "child-centered" recommendations. She saw a distinct similarity in the discourse used by schools, the media (magazines and television programs), the state, and child-development professionals.

Teachers and child-development professionals saw the role of mothers in terms of a necessary collaboration between mothers and schools for the child to succeed not only in school but also in life. Because of unequal resources, all mothers do not participate in this discourse of "good" child development the same way. Griffith found that working-class mothers were perceived as weaker than middle-class mothers in the "stimulation" effort of schooling. Griffith argues that this child-development discourse, embedded in the school institution, perpetuates the reproduction of class by making middle-class ideals for family–school relations the norm for everyone.

Compulsory Heterosexuality

The second illustration of institutional ethnography is taken from Didi Khayatt's (1995) study of the institutionalization of compulsory heterosexuality in schools and its effects on lesbian students. In 1990, Khayatt began her research by interviewing 12 Toronto lesbians, 15 to 24 years of age. Beginning with the young women's viewpoint, she expanded her inquiry to other students, teachers, guidance counselors, and administrators.

Khayatt found that the school's administrative practices generated a *compulsory heterosexuality,* which produced a sense of marginality and vulnerability among lesbian students. For example, the school didn't punish harassment and name-calling directed at gay students. The issue of homosexuality was excluded from the curriculum lest it appear to students as an alternative to heterosexuality.

In both of the studies I've described, the inquiry began with the women's standpoint—mothers and lesbian students. However, instead of emphasizing the subjects' viewpoints, both analyses focused on the power relations that shaped these women's experiences and reality.

Participatory Action Research

Our final field research paradigm takes us further along in our earlier discussion of the status and power relationships linking researchers to the subjects of their research. Within the **participatory action research (PAR)** paradigm, the researcher's function is to serve as a resource to those being studied—typically, disadvantaged groups—as an opportunity for them to act effectively in their own interest. The disadvantaged subjects define their problems, define the remedies desired, and take the lead in designing the research that will help them realize their aims.

This approach began in Third World research development, but it spread quickly to Europe and North America (Gaventa 1991). It comes from a vivid critique of classical social science research. According to the PAR paradigm, traditional research is perceived as an "elitist model" (Whyte, Greenwood, and Lazes 1991) that reduces the "subjects" of research to "objects" of research. According to many advocates of the PAR perspective, the distinction between the researcher and the researched should disappear. They argue that the subjects who will be affected by research should also be responsible for its design.

Implicit in this approach is the belief that research functions not only as a means of knowledge production but also as a "tool for the

participatory action research (PAR) An approach to social research in which the people being studied are given control over the purpose and procedures of the research; intended as a counter to the implicit view that researchers are superior to those they study.

education and development of consciousness as well as mobilization for action" (Gaventa 1991: 121–22). Advocates of participatory action research equate access to information with power and argue that this power has been kept in the hands of the dominant class, sex, ethnicity, or nation. Once people see themselves as researchers, they automatically regain power over knowledge.

Participatory action research poses a special challenge to researchers. On the one hand, participants in the social situation ideally become empowered to frame research relevant to their needs, as they define those needs. At the same time, the researcher brings special skills and insights that nonresearchers lack. So who should be in charge? Andrew Sense (2006: 1) suggests that this decision may have to be made in the moment: "Do I take the 'passenger' position on the bus or do I take the 'driver' seat and be a little more provocative to energise the session[?] My view at this moment is to judge it on the day."

Examples of this approach include research on community power structures, corporate research, and "right-to-know" movements (Whyte, Greenwood, and Lazes 1991). Here are two examples of corporate research that used a PAR approach.

The Corporation

A participatory action research project took place at the Xerox Corporation at the instigation of leaders of both management and the union. Management's goal was to lower costs so that the company could thrive in an increasingly competitive market. The union suggested a somewhat broader scope: improving the quality of working life while lowering manufacturing costs and increasing productivity.

Company managers began by focusing attention on shop-level problems; they were less concerned with labor contracts or problematic managerial policies. At the time, management had a plan to start an "outsourcing" program that would lay off 180 workers, and the union had begun mobilizing to oppose the plan. A consultant hired by Xerox, spent the first month convincing management and the union to create a "cost study team" (CST) that included workers in the wire harness department.

Eight full-time workers were assigned to the CST for six months. Their task was to study

the possibilities of making changes that would save the company $3.2 million and keep the 180 jobs. The team had access to all financial information and was authorized to call on anyone within the company. This strategy allowed workers to make suggestions outside the realm usually available to them. According to Whyte and his colleagues, "reshaping the box enabled the CST to call upon management to explain and justify all staff services" (1991: 27). Because of the changes suggested by the CST and implemented by management, the company saved the targeted $3.2 million.

Management was so pleased by this result that it expanded the wire harness CST project to three other departments that were threatened by competition. Once again, management was happy about the money saved by the teams of workers.

The Xerox case study is an interesting example of participatory action research because it shows how the production of knowledge does not always have to be an elitist enterprise. The "experts" do not necessarily have to be the professionals. According to Whyte and his colleagues, "At Xerox, participatory action research created and guided a powerful process of organizational learning—a process whereby leaders of labor and management learned from each other and from the consultant/facilitator, while he learned from them" (1991: 30).

PAR and Welfare Policy

Participatory action research often involves poor people, as they are typically less able than other groups to influence the policies and actions that affect their own lives. Bernita Quoss, Margaret Cooney, and Terri Longhurst (2000) report a research project involving welfare policy in Wyoming. University students, many of them welfare recipients, undertook research and lobbying efforts aimed at getting Wyoming to accept postsecondary education as "work" under the state's new welfare regulations.

This project began against the backdrop of the 1996 Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA), which

> eliminated education waivers that had been available under the previous welfare law, the 1988 Family Support Act (FSA). These waivers had permitted eligible participants in

the cash assistance AFDC program to attend college as an alternative to work training requirements. Empirical studies of welfare participants who received these waivers have provided evidence that education, in general, is the most effective way to stay out of poverty and achieve self-sufficiency.

(Quoss, Cooney, and Longhurst 2000: 47)

The students began by establishing an organization called Empower and by making presentations on campus to enlist broad student and faculty support. They compiled existing research relevant to the issue and established relationships with members of the state legislature. By the time the 1997 legislative session opened, the students were actively engaged in the process of modifying state welfare laws to offset the shift in federal policy.

The students prepared and distributed fact sheets and other research reports that would be relevant to the legislators' deliberations. They attended committee meetings and lobbied legislators on a one-to-one basis. When erroneous or misleading data were introduced into the discussions, the student-researchers were on hand to point out the errors and offer corrections.

Ultimately, they succeeded. Welfare recipients in Wyoming were allowed to pursue postsecondary education as an effective route out of poverty.

Some researchers speak of **emancipatory research**, which Ardha Danieli and Carol Woodhams (2005: 284) define as "first and foremost a process of producing knowledge which will be of benefit to oppressed people; a political outcome." Both qualitative and quantitative methods can be used to pursue this goal, but it goes well beyond simply learning what's so, even as seen from the subjects' point of view. The authors focus on the study of disability, and they note similarities in the development of emancipatory research and early feminist research.

John Barker and Fiona Smith (2012) offer an extensive review of research using photography to study children and young people. (The March 2012 issue of the *International Journal for Social Research Methodology*, containing this article, is devoted to creative new techniques for studying young subjects.) See the Research in Real Life box, "Pencils and Photos in the Hands of Research Subjects," for a discussion of research using similar innovative methodologies.

As you can see, the seemingly simple process of observing social action as it occurs has subtle though important variations. As we saw in Chapter 2, all our thoughts occur within and are shaped by paradigms, whether we're conscious of it or not. Qualitative field researchers have been unusually deliberate in framing a variety of paradigms to enrich the observation of social life.

The impact of researcher paradigms on the conduct of research is nowhere more explicitly recognized than in the case of *kaupapa Maori research,* a form of participatory action research developed within the indigenous Maori community of New Zealand. As Shayne Walker, Anaru Eketone, and Anita Gibbs (2006) report, an adherence to Maori culture shapes not only the purposes of such research but also its processes and practices. In a study of foster care, for example, the purpose of the study was established by those most directly concerned. The method of collecting data conformed to Maori practices, including public gatherings. The implications derived from the analysis of data were tailored to Maori ways of doing things.

Conducting Qualitative Field Research

So far in this chapter we've examined the kinds of topics appropriate to qualitative field research, special considerations in doing this kind of research, and a sampling of paradigms that direct different types of research efforts. Along the way we've seen some examples that illustrate field research in action. To round out the picture, we turn now to specific ideas and techniques for conducting field research, beginning with how researchers prepare for work in the field.

Preparing for the Field

Suppose for the moment that you've decided to undertake field research on a campus political organization. Let's assume further that you're not a member of that group, that you do not know a great deal about it, and that you'll

emancipatory research Research conducted for the purpose of benefiting disadvantaged groups.

Research in Real Life

Pencils and Photos in the Hands of Research Subjects

How would you go about studying the life conditions of Peruvian Indians living in the Amazon rainforest? With a minimal telecommunications infrastructure and a slow ferry-based postal service in the vast region, a mail or telephone survey wouldn't be the best approach. It might occur to you to conduct in-depth interviews in which you would work from an outline of topics to be covered. Arvind Singhal and Elizabeth Rattine-Flaherty (2006) opted for a very different approach, which put the subjects of study more in control of the research and allowed for important but unexpected discoveries. They derived their inspiration from the work of the renowned Brazilian educator, Paulo Freire, who once set out to measure exploitation among street children. Instead of interviewing them, he gave them cameras and asked them to bring back photographs of exploitation. As Singhal and Rattine-Flaherty report:

> One child took a photo of a nail on a wall. It made no sense to adults, but other children were in strong agreement. The ensuing discussions showed that many young boys of that neighborhood worked in the shoe-shine business. Their clients were mainly in the city, not in the barrio where they lived. As their shoe-shine boxes were too heavy for them to carry, these boys rented a nail on a wall (usually in a shop), where they could hang their boxes for the night. To them, that nail on the wall represented "exploitation." The "nail on the wall" photograph spurred widespread discussions in the Peruvian barrio about other forms of institutionalized exploitation, including ways to overcome them.
>
> *(2006: 314)*

Singhal and Rattine-Flaherty's research involved gauging the quality of life in the Peruvian Amazon and assessing the impact of programs launched by a Peruvian nongovernmental organization (NGO), Minga Peru. To view society through the eyes of children, the researchers set up drawing sessions with colored pencils. In the spirit of reciprocity, one of the authors sketched pictures of snowmen and jack-o'-lanterns that were a part of her childhood in the Midwest. In addition to depicting life in their villages and their close relationship with the natural environment, the children's sketches often featured examples of social change being brought about by development programs of the NGO.

These include sketches of chicken coops, fish farms, and agro-forestry projects. These enterprises, all launched by Minga Peru, began in the Peruvian Amazon only in the past few years. For children to sketch these "new" initiatives in their pictures on their own, without prompts, is noteworthy.

> *(2006: 322)*

The photographs taken by the adult women in the Freire study were equally revealing. Several drew attention to the patriarchal social structure. As the authors report,

> Several photographs depicted the subservient position of the Amazonian women relative to men, a situation that Minga Peru seeks to address. For instance, Adela's picture shows a middle-aged Amazonian woman and her husband sitting on their porch and having a conversation. The woman, sporting a forlorn expression, sits with her legs crossed while her husband stares directly into the camera, squatting with his arms and feet spread in an open position. Especially noticeable is the physical distance of about 10 feet that separates the woman and the man. When Adela was asked why she took the picture and why were the man and woman sitting so far apart, she noted: "The woman is sitting at one side of the house and he is on the other and this was not anything unusual." Upon probing, we learned that Amazonian men determine how close the couple sits. If they are sitting closer, and if the man has his arm around his partner, it is his decision to do so. This authority also applies to initiation of sex: The man determines if and when sex will happen.
>
> *(2006: 323–24)*

This research not only illustrates some unusual data-collection techniques, but it also represents the spirit of emancipatory research. Using similar techniques but with a different end in mind, Pat O'Connor (2006) engaged in participatory action research by asking Irish adolescents to write essays about themselves and about Ireland, including drawings, poems, and songs looking for evidence of the impact of globalization in Ireland. Both studies demonstrate that qualitative field research can be more than just a matter of observing and interviewing.

Sources: Pat O'Connor. 2006. "Globalization, Individualization and Gender in Adolescents' Texts." *International Journal of Social Research Methodology* 9 (4): 261–77; Arvind Singhal and Elizabeth Rattine-Flaherty. 2006. "Pencils and Photos as Tools of Communicative Research and Praxis: Analyzing Minga Peru's Quest for Social Justice in the Amazon." *International Communication Gazette* 68 (4): 313–30.

identify yourself to the participants as a researcher. This section will use this example and others to discuss some of the ways you might prepare yourself before undertaking direct observations.

As is true of all research methods, you would be well advised to begin with a search of the relevant literature, filling in your knowledge of the subject and learning what others have said about it. (Library research is discussed at length in Appendix A.)

In the next phase of your research, you might wish to discuss the student political group with others who have already studied it or with anyone else likely to be familiar with it. In particular, you might find it useful to discuss the group with one or more informants (discussed in Chapter 7). Perhaps you have a friend who is a member, or you can meet someone who is. This aspect of your preparation is likely to be more effective if your relationship with the informant extends beyond your research role. In dealing with members of the group as informants, you should take care that your initial discussions do not compromise or limit later aspects of your research. Keep in mind that the impression you make on the informant, the role you establish for yourself, may carry over into your later effort. For example, creating the initial impression that you may be an undercover FBI agent is unlikely to facilitate later observations of the group.

You should also be wary about the information you get from informants. Although they may have more direct, personal knowledge of the subject under study than you do, what they "know" is probably a mixture of fact and point of view. Members of the political group in our example (as well as members of opposing political groups) would be unlikely to provide completely unbiased information. Before making your first contact with the student group, then, you should already be quite familiar with it, and you should understand its general philosophical context.

There are many ways to establish your initial contact with the people you plan to study. How you do it will depend, in part, on the role you intend to play. Especially if you decide to take on the role of complete participant, you must find a way to develop an identity with the people to be studied. If you wish to study dishwashers in a restaurant, the most direct method would be to get a job as a dishwasher. In the case of the student political group, you might simply join the group.

Many of the social processes appropriate to field research are open enough to make your contact with the people to be studied rather simple and straightforward. If you wish to observe a mass demonstration, just be there. If you wish to observe patterns in jaywalking, hang around busy streets.

Whenever you wish to make more-formal contact with the people, identifying yourself as a researcher, you must establish a rapport with them. You might contact a participant with whom you feel comfortable and gain that person's assistance. In studying a formal group, you might approach the groups' leaders, or you might find that one of your informants can introduce you. (See the Tips and Tools box, "Establishing Rapport," for more details on this.)

Although you'll probably have many options in making your initial contact with the group, realize that your choice can influence your subsequent observations. Suppose, for example, that you're studying a university and begin with high-level administrators. This choice is likely to have a couple of important consequences. First, your initial impressions of the university will be shaped to some extent by the administrators' views, which will differ quite a bit from those of students or faculty. This initial impression may influence the way you observe and interpret events subsequently—especially if you're unaware of the influence.

Second, if the administrators approve of your research project and encourage students and faculty to cooperate with you, the latter groups will probably look on you as somehow aligned with the administration, which can affect what they say to you. For example, faculty members might be reluctant to tell you about plans to organize through the Teamster's Union.

In making direct, formal contact with the people you want to study, you'll be required to give them some explanation of the purpose of your study. Here again, you face an ethical dilemma. Telling them the complete purpose of your research might eliminate their cooperation altogether or significantly affect their behavior. On the other hand, giving only what you believe would be an acceptable explanation may involve outright deception. Your decisions in this and other matters will probably be largely determined by the purpose of your study, the nature of what you're studying, the observations you wish to use, and similar factors, but you must also take ethical considerations into account.

Previous field research offers no fixed rule—methodological or ethical—to follow in this regard. Your appearance as a researcher,

Tips and Tools

Establishing Rapport

In qualitative field research, it's almost always vital that you be able to establish rapport with those you're observing, especially if your observations include in-depth interviews and interactions. **Rapport** might be defined as an open and trusting relationship. But how do you achieve that?

Let's assume that you'll be identifying yourself as a researcher. You'll need to explain your research purpose in a nonthreatening way. Say that you are there to learn about them and understand them, not to judge them or cause them any problems. This will work best if you

1. *Actually have a genuine interest in understanding the people you're observing and can communicate that interest to them.* This gives them a sense of self-worth, which will increase their willingness to open up to you. Pretending to be interested is not the same as really being interested. In fact, if you aren't interested in learning what things look like from the point of view of those you're observing, you might consider another activity and not waste their time and your own.

2. *Be an attentive listener rather than a talker.* You should not remain mute, of course, but you should talk primarily (a) to elicit more information from the other person or (b) to answer questions they may have about you and your research.

3. *Don't argue with your subjects.* While you don't have to agree with any points of view expressed by your subjects, you should never argue with them nor try to change their minds. Keep reminding yourself that your genuine purpose is to understand their world and how it makes sense to them—whether it works for you or not. A little humility may help with this. You'll be able to hear and understand people better if you don't start out feeling superior to them.

4. *Be relaxed and appropriate to the setting.* Some people are more formal or informal than others, and you'll do well to take on their general style or at least find a way to relax with whatever style is most comfortable for them. If you can get them to relax and enjoy the interaction, you'll have achieved the rapport you need. And you'll probably enjoy the interaction yourself.

regardless of your stated purpose, may result in a warm welcome from people who are flattered that a scientist finds them important enough to study. Or, it may result in your being totally ostracized or worse. It probably wouldn't be a good idea, for example, to burst into a meeting of an organized crime syndicate and announce that you're writing a term paper on organized crime.

Qualitative Interviewing

In part, field research is a matter of going where the action is and simply watching and listening. As the baseball legend Yogi Berra said, "You can see a lot just by observing"—provided that you're paying attention. At the same time, as I've already indicated, field research can involve more-active inquiry. Sometimes it's appropriate to ask people questions and record their answers. Your on-the-spot observations of a full-blown riot will lack something if you don't know why people are rioting. Ask somebody.

rapport An open and trusting relationship; especially important in qualitative research between researchers and the people they're observing.

When Cecilia Menjívar (2000) wanted to learn about the experiences of Salvadoran immigrants in San Francisco, she felt in-depth interviews would be a useful technique, along with personal observations. Before she was done, she had discovered a much more complex system of social processes and structures than we would have imagined. Although it was important for new immigrants to have a support structure of family members already in the United States, Menjívar found that her interviewees were often reluctant to call on relatives for help, for several reasons. On the one hand, they might jeopardize those family members who were here illegally and living in poverty. At the same time, asking for help would put them in debt to those helping them out. Menjívar also discovered that Salvadoran gender norms put women immigrants in an especially difficult situation, because they were largely prohibited from seeking the help of men they weren't related to, lest they seem to obligate themselves sexually. These are the kinds of discoveries that can emerge from open-ended, in-depth interviewing.

We've already discussed interviewing in Chapter 9, and much of what was said there applies to qualitative field interviewing. The

interviewing you'll do in connection with field observation, however, is different enough to demand a separate treatment. In surveys, questionnaires are rigidly structured; however, less-structured interviews are more appropriate to field research. Herbert and Riene Rubin (1995: 43) describe the distinction as follows: "Qualitative interviewing design is flexible, iterative, and continuous, rather than prepared in advance and locked in stone." They elaborate in this way:

> Design in qualitative interviewing is iterative. That means that each time you repeat the basic process of gathering information, analyzing it, winnowing it, and testing it, you come closer to a clear and convincing model of the phenomenon you are studying. . . .
>
> The continuous nature of qualitative interviewing means that the questioning is redesigned throughout the project.
>
> *(1995: 46–47)*

Unlike a survey, a **qualitative interview** is an interaction between an interviewer and a respondent in which the interviewer has a general plan of inquiry, including the topics to be covered, but not a set of questions that must be asked with particular words and in a particular order. At the same time, the qualitative interviewer, like the survey interviewer, must be fully familiar with the questions to be asked. This allows the interview to proceed smoothly and naturally.

A qualitative interview is essentially a conversation in which the interviewer establishes a general direction for the conversation and pursues specific topics raised by the respondent. Ideally, the respondent does most of the talking. If you're talking more than 5 percent of the time, that's probably too much.

Steinar Kvale (1996: 3–5) offers two metaphors for interviewing: the interviewer as a "miner" or as a "traveler." The first model assumes that the subject possesses specific information and that the interviewer's job is to dig it out. By contrast, in the second model, the interviewer

> wanders through the landscape and enters into conversations with the people encountered. The traveler explores the many domains of the country, as unknown territory or with maps, roaming freely around the territory. . . . The interviewer

> wanders along with the local inhabitants, asks questions that lead the subjects to tell their own stories of their lived world.

Asking questions and noting answers is a natural human process, and it seems simple enough to add it to your bag of tricks as a field researcher. Be a little cautious, however. Wording questions is a tricky business. All too often, the way we ask questions subtly biases the answers we get. Sometimes we put our respondent under pressure to look good. Sometimes we put the question in a particular context that omits altogether the most relevant answers.

Suppose, for example, that you want to find out why a group of students is rioting and pillaging on campus. You might be tempted to focus your questioning on how students feel about the dean's recent ruling that requires students always to carry *The Practice of Social Research* with them on campus. (Makes sense to me.) Although you may collect a great deal of information about students' attitudes toward the infamous ruling, they may be rioting for some other reason. Perhaps most are simply joining in for the excitement. Properly done, field research interviewing enables you to find out.

In both qualitative and quantitative research, we tend to think of using face-to-face or telephone interviews. When Nicole Ison (2009) set out to conduct in-depth interviews with young people with cerebral palsy, their speech difficulties created a special problem. Her solution was to conduct e-mail interviews. Even in those cases where typing may have been difficult, the subjects could work at their own pace, avoiding the frustration that would probably have attended spoken interviews. Subjects could create their responses and review them to be sure they had accurately expressed their intended communications.

Although you may set out to conduct interviews with a reasonably clear idea of what you want to ask, one of the special strengths of field research is its flexibility. In particular, the answers evoked by your initial questions

---

qualitative interview Contrasted with survey interviewing, the qualitative interview is based on a set of topics to be discussed in depth rather than based on the use of standardized questions.

should shape your subsequent ones. It doesn't work merely to ask preestablished questions and record the answers. Instead, you need to ask a question, listen carefully to the answer, interpret its meaning for your general inquiry, and then frame another question either to dig into the earlier answer or to redirect the person's attention to an area more relevant to your inquiry. In short, you need to be able to listen, think, and talk almost at the same time.

The discussion of probes in Chapter 9 provides a useful guide to getting answers in more depth without biasing later answers. More generally, field interviewers need the skills involved in being a good listener. Be more interested than interesting. Learn to say things like "How is that?" "In what ways?" "How do you mean that?" "What would be an example of that?" Learn to look and listen expectantly, and let the person you're interviewing fill in the silence.

At the same time, you can't afford to be a totally passive receiver. You'll go into your interviews with some general (or specific) questions you want answered and some topics you want addressed. At times you'll need the skill of subtly directing the flow of conversation.

There's something we can learn in this regard from the martial arts. The aikido master never resists an opponent's blow but instead accepts it, joins with it, and then subtly redirects it in a more appropriate direction. Field interviewing requires an analogous skill. Instead of trying to halt your respondent's line of discussion, learn to take what he or she has just said and branch that comment back in the direction appropriate to your purposes. Most people love to talk to anyone who's really interested. Stopping their line of conversation tells them that you are not interested; asking them to elaborate in a particular direction tells them that you are.

Consider this hypothetical example in which you're interested in why college students chose their majors.

YOU: What are you majoring in?

RESP: Engineering.

YOU: I see. How did you come to choose engineering?

RESP: I have an uncle who was voted the best engineer in Arizona in 2005.

YOU: Gee, that's great.

RESP: Yeah. He was the engineer in charge of developing the new civic center in Tucson. It was written up in most of the engineering journals.

YOU: I see. Did you talk to him about your becoming an engineer?

RESP: Yeah. He said that he got into engineering by accident. He needed a job when he graduated from high school, so he went to work as a laborer on a construction job. He spent eight years working his way up from the bottom, until he decided to go to college and come back nearer the top.

YOU: So is your main interest civil engineering, like your uncle, or are you more interested in some other branch of engineering?

RESP: Actually, I'm leaning more toward electrical engineering—computers, in particular. I started messing around with a Macintosh when I was in high school, and my long-term plan is . . .

Notice how the interview begins by wandering off into a story about the respondent's uncle. The first attempt to focus things back on the student's own choice of major ("Did you talk to your uncle . . . ?") fails. The second attempt ("So is your main interest . . . ?") succeeds. Now the student is providing the kind of information you're looking for. It's important for field researchers to develop the ability to "control" conversations in this fashion. At the same time, of course, you need to be on the alert for "distractions" that point to unexpectedly important aspects of your research interest.

Herbert and Riene Rubin offer several ways to control a "guided conversation," including the following:

If you can limit the number of main topics, it is easier to maintain a conversational flow from one topic to another. Transitions should be smooth and logical. "We have been talking about mothers, now let's talk about fathers," sounds abrupt. A smoother transition might be, "You mentioned your mother did not care how you performed in school— was your father more involved?" The more abrupt the transition, the more it sounds like

the interviewer has an agenda that he or she wants to get through, rather than wanting to hear what the interviewee has to say.

(1995: 123)

Because field research interviewing is so much like normal conversation, researchers must keep reminding themselves that they are not having a normal conversation. In normal conversations, each of us wants to come across as an interesting, worthwhile person. If you watch yourself the next time you chat with someone you don't know too well, you'll probably find that much of your attention is spent on thinking up interesting things to say—contributions to the conversation that will make a good impression. Often, we don't really hear each other, because we're not really listening—we're too busy thinking of what we'll say next. As an interviewer, the desire to appear interesting is counterproductive. The interviewer needs to make the other person seem interesting, by being interested—and by listening more than talking. (Do this in ordinary conversations, and people will actually regard you as a great conversationalist.)

John Lofland and his colleagues (2006: 69–70) suggest that researchers should adopt the role of the "socially acceptable incompetent" when interviewing. That is, offer yourself as someone who does not understand the situation you find yourself in and must be helped to grasp even the most basic and obvious aspects of that situation: "A naturalistic investigator, almost by definition, is one who does not understand. She or he is 'ignorant' and needs to be 'taught.' This role of watcher and asker of questions is the quintessential *student* role" (Lofland et al. 2006: 69).

Interviewing needs to be an integral part of the entire field research process. Later, I'll stress the need to review your observational notes every night—making sense out of what you've observed, getting a clearer feel for the situation you're studying, and finding out what you should pay more attention to in further observations. In the same fashion, you'll need to review your notes on interviews, recording especially effective questions and detecting all those questions you should have asked but didn't. Start asking such questions the next time you interview. If you've recorded the interviews, replay them as a useful preparation for future interviews.

Steinar Kvale (1996: 88) details seven stages in the complete interviewing process:

1. *Thematizing:* Clarifying the purpose of the interviews and the concepts to be explored

2. *Designing:* Laying out the process through which you'll accomplish your purpose, including a consideration of the ethical dimension

3. *Interviewing:* Doing the actual interviews

4. *Transcribing:* Creating a written text of the interviews

5. *Analyzing:* Determining the meaning of gathered materials in relation to the purpose of the study

6. *Verifying:* Checking the reliability and validity of the materials

7. *Reporting:* Telling others what you've learned

As with all other aspects of field research, interviewing improves with practice. Fortunately, it's something you can practice any time you want. Practice on your friends.

Focus Groups

Although our discussions of field research so far have focused on studying people in the process of living their lives, researchers sometimes bring people into the laboratory for qualitative interviewing and observation. The focus group method, which is also called group interviewing, is essentially a qualitative method. It is based on structured, semistructured, or unstructured interviews. It allows the researcher/interviewer to question several individuals systematically and simultaneously. This data-collection technique is frequently used in political and market research but is used for other purposes as well. In *Silent Racism,* for example, Barbara Trepagnier (2006) used focus groups to examine the persistence of racism among "well-meaning white people."

In a hypothetical market-research example, imagine that you're thinking about introducing a new product. Let's suppose that you've invented a new computer that not only does word processing, spreadsheets, data analysis, and the like but also contains a fax machine, AM/FM/TV tuner, GPS, MP3, climate change calculator, automobile diagnostic system, microwave oven, denture cleaner, and coffeemaker. To highlight its computing and coffee-making features,

you're thinking of calling it "The Compulator." You figure the new computer will sell for about $28,000, and you want to know whether people are likely to buy it. Your prospects might be well served by focus groups.

In a **focus group**, typically 5 to 15 people are brought together in a private, comfortable environment to engage in a guided discussion of some topic—in this case, the acceptability and salability of The Compulator. The subjects are selected on the basis of relevance to the topic under study. Given the likely cost of The Compulator, your focus group participants would probably be limited to upper-income groups, for example. Other, similar considerations might figure into the selection.

Participants in focus groups are not likely to be chosen through rigorous probability-sampling methods. This means that the participants do not statistically represent any meaningful population. However, the purpose of the study is to explore rather than to describe or explain in any definitive sense. Nevertheless, typically more than one focus group is convened in a given study because of the serious danger that a single group of 7 to 12 people will be too atypical to offer any generalizable insights.

William Gamson (1992) used focus groups to examine how U.S. citizens frame their views of political issues. Having picked four issues—affirmative action, nuclear power, troubled industries, and the Arab–Israeli conflict—Gamson undertook a content analysis of press coverage to get an idea of the media context within which we think and talk about politics. Then the focus groups were convened for a firsthand observation of the process of people discussing issues with their friends.

Richard Krueger (1988: 47) points to five advantages of focus groups:

1. The technique is a socially oriented research method capturing real-life data in a social environment.
2. It has flexibility.

---

focus group A group of subjects interviewed together, prompting a discussion. The technique is frequently used by market researchers, who ask a group of consumers to evaluate a product or discuss a type of commodity, for example.

---

3. It has high face validity.
4. It has speedy results.
5. It is low in cost.

(1988: 47)

In addition to these advantages, group dynamics frequently bring out aspects of the topic that would not have been anticipated by the researcher and would not have emerged from interviews with individuals. In a side conversation, for example, a couple of the participants might start joking about the results of leaving out one letter from a product's name. This realization might save the manufacturer great embarrassment later on.

Krueger (1988: 44–45) also notes some disadvantages of the focus group method, however:

1. Focus groups afford the researcher less control than individual interviews.
2. Data are difficult to analyze.
3. Moderators require special skills.
4. Difference between groups can be troublesome.
5. Groups are difficult to assemble.
6. The discussion must be conducted in a conducive environment.

(1988: 48)

As we've seen, the group interview presents several advantages, but it also has its challenges. In a focus group interview, much more than in any other type of interview, the interviewer has to develop the skills of a moderator. Controlling the dynamic within the group is a major challenge. Letting one person dominate the focus group interview reduces the likelihood that the other subjects will express themselves. This can generate the problem of group conformity or groupthink, which is the tendency for people in a group to conform with the opinions and decisions of the most outspoken members of the group. This danger is compounded by the possibility that only one or two people sometimes dominate the conversation. Interviewers need to be aware of this phenomenon and try to get everyone to participate fully on all the issues brought in the interview. Adding to the challenge, of course, is that the interviewer must resist overdirecting the interview and the interviewees, thus bringing her or his own views into play.

Although focus group research differs from other forms of qualitative field research, it further illustrates the possibilities for doing social research face-to-face with those we wish to understand. In addition, David Morgan (1993) suggests that focus groups are an excellent device for generating questionnaire items for a subsequent survey.

Because they center on a particular topic and take relatively little time, focus groups are typically regarded as an "in-depth" research technique. However, Carolina Överlien, Karin Aronsson, and Margareta Hydén (2005) have used the technique successfully for extended discussions of sexuality, among Swedish teenagers in a youth detention home.

Like other social research techniques, focus groups are adapting to new communication modalities. George Silverman (2005), for example, offers a discussion of telephone and online focus groups.

Recording Observations

The greatest advantage of the field research method is the presence of an observing, thinking researcher on the scene of the action. Even audio recorders and cameras cannot capture all the relevant aspects of social processes, although both of those devices can be quite useful to the field researcher. Consequently, in both direct observation and interviewing, it's vital to make full and accurate notes of what goes on. If possible, take notes on your observations as you observe. When that's not feasible, write down your notes as soon as possible afterward.

In your notes, include both your empirical observations and your interpretations of them. In other words, record what you "know" has happened and what you "think" has happened. Be sure to identify these different kinds of notes for what they are. For example, you might note that Person X spoke out in opposition to a proposal made by a group leader (an observation), that you think this represents an attempt by Person X to take over leadership of the group (an interpretation), and that you think you heard the leader comment to that effect in response to the opposition (a tentative observation).

Of course, you can't observe everything; nor can you record everything you do observe. Just

as your observations will represent a sample of all possible observations, your notes will represent a sample of your observations. The idea, of course, is to record the most pertinent ones.

The Tips and Tools box, "Interview Transcript Annotated with Researcher Memos," provides an extract from an in-depth interview with a woman film director, conducted by Sandrine Zerbib. Notice that the illustration contains a portion of an extensive interview along with some of Zerbib's memos, written during her review of her interview notes later on.

Some of your most important observations can be anticipated before you begin the study; others will become apparent as your observations progress. Sometimes you can make note taking easier by preparing standardized recording forms in advance. In a study of jaywalking, for example, you might anticipate the characteristics of pedestrians that are most likely to be useful for analysis—age, gender, social class, ethnicity, and so forth—and prepare a form in which observations of these variables can be recorded easily. Alternatively, you might develop a symbolic shorthand in advance to speed up recording. For studying audience participation at a mass meeting, you might want to construct a numbered grid representing the different sections of the meeting room; then you could record the location of participants easily, quickly, and accurately.

None of this advance preparation should limit your recording of unanticipated events and aspects of the situation. Quite the contrary, speedy handling of anticipated observations can give you more freedom to observe the unanticipated.

You're already familiar with the process of taking notes, just as you already have at least informal experience with field research in general. Like good field research, however, good note taking requires careful and deliberate attention and involves specific skills. Some guidelines follow. (You can learn even more from Lofland et al. 2006: 110–17.)

First, don't trust your memory any more than you have to—it's untrustworthy. To illustrate this point, try this experiment. Recall the last three or four movies you saw that you really liked. Now, name five of the actors or actresses. Who had the longest hair? Or can you remember

Tips and Tools

Interview Transcript Annotated with Researcher Memos

Thursday August 26, 12:00–1:00

R is the interviewer and J is the director-subject. Numbers in brackets represent a word that was inaudible from the interview. Each number corresponds to the number that appeared on the transcribing machine counter, with each interview starting at zero. The numbers help the researcher locate a passage quickly when he or she reviews the interview.

R: What is challenging for women directors on a daily experience, on a daily life?

J: Surviving.

R: OK. Could you develop a little bit on that? [I need to work on my interview schedule so that my interviewee answers with more elaboration without having to probe.]

J: Yeah, I mean it's all about trying to get, you know, in, trying to get the job, and try, you know, to do a great job so that you are invited back to the next thing. And particularly since they are so many, you know, difficulties in women directing. It makes it twice as hard to gain into this position where you do an incredible job, because . . . you can't just do an average job, you have to [347] do this job that just knocks your socks off all the time, and sometimes you don't get the opportunity to do that, because either you don't have a good producer or you have so many pressures that you can't see straight or your script is lousy, and you have to make a silk purse out of sow's hair. You know, you have a lot of extra strikes against you than the average guy who has similar problems, because you are a woman and they look at it, and women are more visible than men . . . in unique positions.

[It seems that Joy is talking about the particularities of the film industry. There are not that many opportunities and in order to keep working, she needs to build a certain reputation. It is only by continuing to direct that she can maintain or improve her reputation. She thinks that it is even harder for women but does not explain it.]

R: Hum . . . what about on the set did you experience, did it feel . . . did people make it clear that you were a woman, and you felt treated differently? [I am trying to get her to speak about more specific and more personal experiences without leading her answer.]

J: Yeah, oh yeah, I mean . . . a lot of women have commiserated about, you know when you have to walk on the set for the first time, they're all used to working like a well-oiled machine and they say, "Oh, here is the woman, something different" and sometimes they can be horrible, they can resist your directing and they can, they can sabotage you, by taking a long time to light, or to move sets, or to do something . . . and during that time you're wasting time, and that goes on a report, and the report goes to the front [368] office, and, you know, and so on and so on and so on and so forth. And people upstairs don't know what the circumstances are, and they are not about to fire a cinematographer that is on their show for ever and ever . . . nor do they want to know that this guy is a real bastard, and making your life a horror. They don't want to know that, so therefore, they go off, because she's a woman let's not hire any more women, since he has problems with women. You know, so, there is that aspect.

[I need to review the literature on institutional discrimination. It seems that the challenges that Joy is facing are not a matter of a particular individual. She is in a double bind situation where whether she complains or not, she will not be treated equal to men. Time seems to be one quantifiable measurement of how well she does her job and, as observed in other professions, the fact that she is a woman is perceived as a handicap. Review literature on women in high management positions. I need to keep asking about the dynamics between my interviewees and the crewmembers on the set. The cinematographer has the highest status on the set under the director. Explore other interviews about reasons for conflict between them.]

[Methods (note to myself for the next interviews): Try to avoid phone interviews unless specific request from the interviewee. It is difficult to assess how the interviewee feels with the questions. Need body language because I become more nervous about the interview process.]

what your boyfriend, girlfriend, or best friend was wearing yesterday? (Remembering what *you* were wearing yesterday may even be a challenge.)

Even if you pride yourself on having a photographic memory, it's a good idea to take notes either during the observation or as soon afterward as possible. If you take notes during observation, do it unobtrusively, because people are likely to behave differently if they see you taking down everything they say or do.

Second, it's usually a good idea to take notes in stages. In the first stage, you may need to take sketchy notes (words and phrases) in order to keep abreast of what's happening. Then go off by yourself and rewrite your notes in more detail. If you do this soon after the events you've observed, the sketchy notes should allow you to

recall most of the details. The longer you delay, the less likely you'll be able to recall things accurately and fully.

In his study of bike messengers in New York City, mentioned earlier, Jeffrey Kidder reports on this process:

> I obtained the vast majority of data for this article through informal interviews. I unobtrusively took notes throughout the day and at social events. Upon returning home, these data were compiled into my field notes. During the workday and during races, parties, and other social gatherings, casual conversations provided the truest glimpses into messenger beliefs, ideologies, and opinions. To this end, I avoided formal interviews and instead allowed my questions to be answered by normal talk within the social world.
>
> *(2005: 349)*

I know this method sounds logical, but it takes self-discipline to put it into practice. Careful observation and note taking can be tiring, especially if it involves excitement or tension and if it extends over a long period. If you've just spent eight hours observing and making notes on how people have been coping with a disastrous flood, your first desire afterward will likely be to get some sleep, dry clothes, or get a bite to eat. You may need to take some inspiration from newspaper reporters who undergo the same sorts of hardships then write their stories to meet their deadlines.

Third, you'll inevitably wonder how much you should record. Is it really worth the effort to write out all the details you can recall right after the observational session? The general guideline is yes. Generally, in field research you can't be really sure of what's important and what's unimportant until you've had a chance to review and analyze a great volume of information, so you should record even things that don't seem important at the outset. They may turn out to be significant after all. Also, the act of recording the details of something "unimportant" may jog your memory on something that is important.

Realize that most of your field notes will not be reflected in your final report on the project. Put more harshly, most of your notes will be "wasted." But take heart: Even the richest gold ore yields only about 30 grams of gold per metric ton, meaning that 99.997 percent of the ore is wasted. Yet, that 30 grams of gold can be hammered out to cover an area 18 feet square—the equivalent of about 685 book pages. So take a ton of notes, and plan to select and use only the gold.

Like other aspects of field research (and all research for that matter), proficiency comes with practice. The nice thing about field research is you can begin practicing now and can continue practicing in almost any situation. You don't have to be engaged in an organized research project to practice observation and recording. You might start by volunteering to take the minutes at committee meetings, for example. Or just pick a sunny day on campus, find a shady spot, and try observing and recording some specific characteristics of the people who pass by. You can do the same thing at a shopping mall or on a busy street corner. Remember that observing and recording are professional skills and, like all worthwhile skills, they improve with practice.

Strengths and Weaknesses of Qualitative Field Research

Like all research methods, qualitative field research has distinctive strengths and weaknesses. As I've already indicated, field research is especially effective for studying subtle nuances in attitudes and behaviors and for examining social processes over time. As such, the chief strength of this method lies in the depth of understanding it permits. Whereas other research methods may be challenged as "superficial," this charge is seldom lodged against field research.

Flexibility is another advantage of field research. As discussed earlier, you can modify your field research design at any time. Moreover, you're always prepared to engage in field research, whenever the occasion should arise, whereas you could not as easily initiate a survey or an experiment.

Field research can be relatively inexpensive as well. Other social science research methods may require costly equipment or an expensive research staff, but field research typically can be undertaken by one researcher with a notebook and a pencil. This is not to say that field research is never expensive. The nature of the research

project may require a large number of trained observers, for example. Expensive recording equipment may be needed. Or you may wish to undertake participant observation of interactions in pricey Paris nightclubs.

Field research has several weaknesses as well. First, being qualitative rather than quantitative, it's not an appropriate means for arriving at statistical descriptions of a large population. Observing casual political discussions in Laundromats, for example, would not yield trustworthy estimates of the future voting behavior of the total electorate. Nevertheless, the study could provide important insights into how political attitudes are formed.

To assess field research further, let's focus on the issues of validity and reliability. Recall that validity and reliability are both qualities of measurements. Validity concerns whether measurements actually measure what they're intended to rather than something else. Reliability, on the other hand, is a matter of dependability: If you made the same measurement again and again, would you get the same result? Let's see how field research stacks up in these respects.

Validity

Field research seems to provide measures with greater validity than do survey and experimental measurements, which are often criticized as superficial and not really valid. Let's review a couple of field research examples to see why this is so.

"Being there" is a powerful technique for gaining insights into the nature of human affairs in all their rich complexity. Listen, for example, to what this nurse reports about the impediments to patients' coping with cancer:

> Common fears that may impede the coping process for the person with cancer can include the following:
>
> —Fear of death—for the patient, and the implications his or her death will have for significant others.
>
> —Fear of incapacitation—because cancer can be a chronic disease with acute episodes that may result in periodic stressful periods, the variability of the person's ability to cope and constantly adjust may require a dependency upon others for activities of daily living and may consequently become a burden.

> —Fear of alienation—from significant others and health care givers, thereby creating helplessness and hopelessness.
>
> —Fear of contagion—that cancer is transmissible and/or inherited.
>
> —Fear of losing one's dignity—losing control of all bodily functions and being totally vulnerable.
>
> *(Garant 1980: 2167)*

Observations and conceptualizations such as these are valuable in their own right. In addition, they can provide the basis for further research—both qualitative and quantitative.

Now listen to what Joseph Howell has to say about "toughness" as a fundamental ingredient of life on Clay Street, a white, working-class neighborhood in Washington, D.C.:

> Most of the people on Clay Street saw themselves as fighters in both the figurative and literal sense. They considered themselves strong, independent people who would not let themselves be pushed around. For Bobbi, being a fighter meant battling the welfare department and cussing out social workers and doctors upon occasion. It meant spiking Barry's beer with sleeping pills and bashing him over the head with a broom. For Barry it meant telling off his boss and refusing to hang the door, an act that led to his being fired. It meant going through the ritual of a duel with Al. It meant pushing Bubba around and at times getting rough with Bobbi.
>
> June and Sam had less to fight about, though if pressed they both hinted that they, too, would fight. Being a fighter led Ted into near conflict with Peg's brothers, Les into conflict with Lonnie, Arlene into conflict with Phyllis at the bowling alley, etc.
>
> *(1973: 292)*

Even without having heard the episodes Howell refers to in this passage, you have the distinct impression that Clay Street is a tough place to live. That "toughness" shows far more powerfully through these field observations than it would in a set of statistics on the median number of fistfights occurring during a specified period.

These examples point to the superior validity of field research, as compared with surveys and experiments. The kinds of comprehensive

measurements available to the field researcher tap a depth of meaning in concepts such as common fears of cancer patients and "toughness" (or concepts such as liberal and conservative) that are generally unavailable to surveys and experiments. Instead of specifying concepts, field researchers commonly give detailed illustrations.

Reliability

Field research, however, can pose problems of reliability. Suppose you were to characterize your best friend's political orientations according to everything you know about him or her. Your assessment of your friend's politics would appear to have considerable validity; certainly it's unlikely to be superficial. We couldn't be sure, however, that another observer would characterize your friend's politics the same way you did, even with the same amount of observation.

Although they are in-depth, field research measurements are also often very personal. How I judge your friend's political orientation depends greatly on my own, just as your judgment depends on your political orientation. Conceivably, then, you could describe your friend as middle-of-the-road, although I might feel that I've been observing a fire-breathing radical.

As I suggested earlier, researchers who use qualitative techniques are conscious of this issue and take pains to address it. Individual researchers often sort out their own biases and points of view, and the communal nature of science means that their colleagues will help them in that regard. Nevertheless, it's prudent to be wary of purely descriptive measurements in field research—your own, or someone else's. If a researcher reports that the members of a club are fairly conservative, such a judgment is unavoidably linked to the researcher's own politics. You can be more trusting of comparative evaluations: identifying who is more conservative than who, for example. Even if you and I had different political orientations, we would probably agree pretty much in ranking the relative conservatism of the members of a group.

As a means for both increasing and documenting the trustworthiness of qualitative research, Glenn Bowen (2009) illustrates the use of an "audit trail," which records the researcher's decisions throughout the conduct of the research and the analysis of data. Decisions on the coding of interview responses would be an example. Some computer programs for qualitative data analysis provide for the recording of an audit trail.

While the audit trail is suggested to counter concerns that qualitative analysis might lack rigor, a similar technique would be appropriate for quantitative research. While the results of measurement decisions in designing a quantitative survey are explicit in the actual wording of questionnaires, the reasoning behind those decisions is not always obvious.

As we've seen, field research is a potentially powerful tool for social scientists, one that provides a useful balance to the strengths and weaknesses of experiments and surveys. The remaining chapters of Part 3 present additional modes of observation available to social researchers.

Ethics and Qualitative Field Research

As I've noted repeatedly, all forms of social research raise ethical issues. By bringing researchers into direct and often intimate contact with their subjects, field research raises ethical concerns in a particularly dramatic way. Here are some of the issues mentioned by John and Lyn Lofland (1995: 63):

- Is it ethical to talk to people when they do not know you will be recording their words?
- Is it ethical to get information for your own purposes from people you hate?
- Is it ethical to see a severe need for help and not respond to it directly?
- Is it ethical to be in a setting or situation but not commit yourself wholeheartedly to it?
- Is it ethical to develop a calculated stance toward other humans, that is, to be strategic in your relations?
- Is it ethical to take sides or to avoid taking sides in a factionalized situation?
- Is it ethical to "pay" people with trade-offs for access to their lives and minds?
- Is it ethical to "use" people as allies or informants in order to gain entree to other people or to elusive understandings?

Participation observation brings special ethical concerns with it. When you ask people to reveal their inner thoughts and actions to you, you may be opening them up to a degree of suffering: perhaps recalling troubling experiences, for example, as in the earlier example of interviewing cancer patients. Moreover, you are also asking them to risk the public disclosure of what they have confided in you, and you are strictly obligated to honor their confidences. We have seen cases of researchers going to jail rather than reveal the private matters they observed in confidence.

Geoff Pearson (2009) examines the sticky question of how participant observers should behave when studying people routinely engaged in criminal activities. The researcher's refusal to join in such illegal behavior might very well alter what is being studied and, in some cases, risk the researcher's study and/or safety. On the other hand, are researchers justified in breaking the law in such cases? Obviously the severity of the crimes would affect your decisions, but when you examine such ethical questions in depth, you are likely to find yourself entering numerous gray areas. Planning and conducting field research in a responsible way requires attending to these and other ethical concerns.

MAIN POINTS

Introduction

- Field research involves the direct observation of social phenomena in their natural settings. Typically, field research is qualitative rather than quantitative.
- In field research, observation, data processing, and analysis are interwoven, cyclical processes.

Topics Appropriate for Field Research

- Field research is especially appropriate for topics and processes that are not easily quantifiable, that are best studied in natural settings, or that change over time. Among these topics are practices, episodes, encounters, roles, relationships, groups, organizations, settlements, social worlds, and lifestyles or subcultures.

Special Considerations in Qualitative Field Research

- Among the special considerations involved in field research are the various possible roles of the observer and the researcher's relationships with subjects. As a field researcher, you must decide whether to observe as an outsider or as a participant, whether or not to identify yourself as a researcher, and how to negotiate your relationships with subjects.

Some Qualitative Field Research Paradigms

- Field research can be guided by any one of several paradigms, such as naturalism, ethnomethodology, grounded theory, case studies and the extended case method, institutional ethnography, and participatory action research.

Conducting Qualitative Field Research

- Preparing for the field involves doing background research, determining how to make contact with subjects, and resolving issues of what your relationship to your subjects will be.
- Field researchers often conduct in-depth interviews that are much less structured than those conducted in survey research. Qualitative interviewing is more of a guided conversation than a search for specific information. Effective interviewing involves skills of active listening and the ability to direct conversations unobtrusively.
- To create a focus group, researchers bring subjects together and observe their interactions as they explore a specific topic.
- Whenever possible, field observations should be recorded as they are made; otherwise, they should be recorded as soon afterward as possible.

Strengths and Weaknesses of Qualitative Field Research

- Among the advantages of field research are the depth of understanding it can provide, its flexibility, and (usually) its low cost.
- Compared with surveys and experiments, field research measurements generally have more validity but less reliability. Also, field research is generally not appropriate for arriving at statistical descriptions of large populations.
- An audit trail records the researcher's decisions throughout the conduct of the research and the analysis of data.

Ethics and Qualitative Field Research

- Conducting field research responsibly involves confronting several ethical issues that arise from the researcher's direct contact with subjects.

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.

case study

emancipatory research

ethnography

ethnomethodology

extended case method

focus group

grounded theory

institutional ethnography

naturalism

participatory action
    research (PAR)

qualitative interview

rapport

reactivity

questions that must be addressed in each interview and others that will be pursued only if appropriate?

Compared with experiments and surveys, field research allows more flexibility as to the timing of the research. Depending on how things go, you may find yourself concluding earlier or later than you had planned. Nevertheless, you should say something in the proposal regarding the schedule you are planning.

PROPOSING SOCIAL RESEARCH: QUALITATIVE FIELD RESEARCH

This chapter has laid out a large number of different possibilities for conducting field research. If you're doing field research, you should indicate the kind of study you plan to do. Will you be the sole observer in the study? If not, how will you select and train the other observers?

Will you be a participant in the events you are observing and, if so, will you identify yourself as a researcher to those you are observing? You might say something about how these choices may affect what you observe, as well as discussing the ethical issues involved.

In earlier exercises, you dealt with the variables you'll examine and the ways you'll select informants and/or people to observe, as well as the times and places for your observations. As this chapter has demonstrated, there are other logistical issues to be worked out. It may be appropriate to describe your note-taking plans if that's likely to be difficult (for example, if you're a participant not identified as a researcher).

If you'll be conducting in-depth interviews, you should include an outline of the topics to be covered in those interviews. Are there topics or
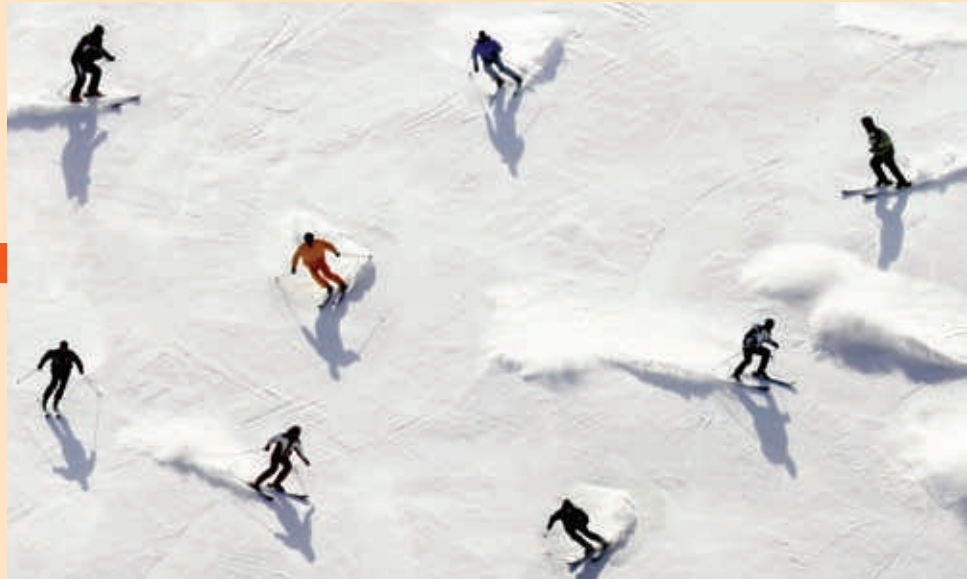
REVIEW QUESTIONS AND EXERCISES

1. Think of some group or activity you participate in or are very familiar with. In two or three paragraphs, describe how an outsider might effectively go about studying that group or activity. What should he or she read, what contacts should be made, and so on?

2. Choose any two of the paradigms discussed in this chapter. Then describe how your hypothetical study from Exercise 1 might be conducted if you followed each. Compare and contrast the way these paradigms might work in the context of your study.

3. To explore the strengths and weaknesses of experiments, surveys, and field research, choose a general research area (such as prejudice, political orientation, education) and write brief descriptions of studies in that area that could be conducted using each of these three methods. In each case, explain why the chosen method is the most appropriate for the study you describe.

4. Return to the example you devised in response to Exercise 1 and list five ethical issues that you can imagine having to confront if you were to undertake your study.

Unobtrusive Research

CHAPTER OVERVIEW

This chapter presents overviews of three unobtrusive research methods: content analysis, the analysis of existing statistics, and comparative and historical research. Each of these methods allows researchers to study social life from afar, without influencing it in the process.

Introduction

With the exception of the complete observer in field research, each of the modes of observation discussed so far requires the researcher to intrude, to some degree, on whatever he or she is studying. This is most obvious in the case of experiments, followed closely by survey research. Even the field researcher, as we've seen, can change things in the process of studying them.

At least one previous example in this book, however, was totally exempt from that danger. Durkheim's analysis of suicide did nothing to affect suicides one way or the other (see Chapter 5). His study is an example of **unobtrusive research**, or methods of studying social behavior without affecting it. As you'll see, unobtrusive measures can be qualitative or quantitative.

This chapter examines three types of unobtrusive research methods: content analysis, analysis of existing statistics, and comparative and historical research. In content analysis, researchers examine a class of social artifacts that usually are written documents such as newspaper editorials. Next, the Durkheim study is an example of the analysis of existing statistics. As you'll see, there are great masses of data all around you, awaiting your use in the understanding of social life. Finally, comparative and historical research, a form of research with a venerable history in the social sciences, is currently enjoying a resurgence of popularity. Like field research, comparative and historical research is usually a qualitative method, one in which the main resources for observation and analysis are historical records. The method's name includes the word *comparative* because social scientists—in contrast to historians who may simply describe a particular set of events—seek to discover common patterns that recur in different times and places.

To set the stage for our examination of these three research methods, I want to draw your attention to an excellent book that should sharpen your senses about the potential for unobtrusive measures in general. It is, among other things, the book from which I take the term *unobtrusive measures.*

In 1966, Eugene Webb and three colleagues published an ingenious little book on social research (revised in 2000) that has become a classic. It focuses on the idea of unobtrusive or nonreactive research. Webb and his colleagues have played freely with the task of learning about human behavior by observing what people inadvertently leave behind them. Do you want to know what exhibits are the most popular at a museum? You could conduct a poll, but people might tell you what they thought you wanted to hear or what might make them look intellectual and serious. You could stand by different exhibits and count the viewers that came by, but people might come over to see what you were doing. Webb and his colleagues suggest that you check the wear and tear on the floor in front of various exhibits. Those that have the most-worn tiles are probably the most popular. Want to know which exhibits are popular with little kids? Look for mucus on the glass cases. To get a sense of the most popular radio stations, you could arrange with an auto mechanic to check what radio stations are programmed in for cars brought in for repair.

The possibilities are limitless. Like a detective investigating a crime, the social researcher looks for clues. If you stop to notice, you'll find that clues of social behavior are all around you. In a sense, everything you see represents the answer to some important social science question—all you have to do is think of the question.

Although problems of validity and reliability crop up in unobtrusive measures, a little ingenuity can either handle them or put them in perspective.

Content Analysis

As I mentioned in the chapter introduction, **content analysis** is the study of recorded human communications. Among the forms suitable for

---

unobtrusive research Methods of studying social behavior without affecting it. Such methods can be qualitative or quantitative.

content analysis The study of recorded human communications, such as books, websites, paintings, and laws.

study are books, magazines, web pages, poems, newspapers, songs, paintings, speeches, letters, e-mail messages, bulletin-board postings on the Internet, laws, and constitutions, as well as any components or collections thereof. Shulamit Reinharz points out that feminist researchers have used content analysis to study "children's books, fairy tales, billboards, feminist nonfiction and fiction books, children's art work, fashion, fat-letter postcards, Girl Scout Handbooks, works of fine art, newspaper rhetoric, clinical records, research publications, introductory sociology textbooks, and citations, to mention only a few" (1992: 146–47). In another example, when William Mirola set out to discover the role of religion in the movements to establish the eight-hour working day in America, his data were taken "from Chicago's labor, religious, and secular presses, from pamphlets, and from speeches given by eight-hour proponents from three representative factions within the movement" (2003: 273).

Topics Appropriate for Content Analysis

Content analysis is particularly well suited to the study of communications and to answering the classic question of communications research: "Who says what, to whom, why, how, and with what effect?" Are popular French novels more concerned with love than novels in the United States are? Was the popular British music of the 1960s more politically cynical than the popular German music during that period? Do political candidates who primarily speak to "bread and butter" issues get elected more often than those who address issues of high principle? Each of these questions suggests a social science research topic: The first might address national character, the second political orientations, and the third political process. Although you might study such topics by observing individual people, content analysis provides another approach.

An early example of content analysis is the work of Ida B. Wells. In 1891, Wells, whose parents had been slaves, wanted to test the widely held assumption that African American men were being lynched in the South primarily for raping white women. As a research method, she examined newspaper articles on

the 728 lynchings reported during the previous ten years. In only a third of the cases were the lynching victims even accused of rape, much less proved guilty. Primarily, they were charged with being insolent, not staying in "their place" (cited in Reinharz 1992: 146).

More recently, the best-selling book *Megatrends 2000* (Naisbitt and Aburdene 1990) used content analysis to determine the major trends in modern U.S. life. The authors regularly monitored thousands of local newspapers a month in order to discover local and regional trends for publication in a series of quarterly reports. Their book examines some of the trends they observed in the nation at large. In a follow-up book (Aburdene 2005), this kind of analysis pointed to such trends as "the power of spirituality" and "the rise of conscious capitalism."

Some topics are more appropriately addressed by content analysis than by any other method of inquiry. Suppose that you're interested in violence on television. Maybe you suspect that the manufacturers of men's products are more likely to sponsor violent TV shows than are sponsors of other products or services. Content analysis would be the best way of finding out.

Briefly, here's what you'd do. First, you'd develop operational definitions of the two key variables in your inquiry: *men's products* and *violence*. The section on coding, later in this chapter, will discuss some of the ways you could do that. Ultimately, you'd need a plan that would allow you to watch TV, classify sponsors, and rate the degree of violence on particular shows.

Next, you'd have to decide what to watch. Probably you'd decide (1) what stations to watch, (2) for what period, and (3) at what hours. Then, you'd stock up on beer and potato chips and start watching, classifying, and recording. Once you'd completed your observations, you'd be able to analyze the data you collected and determine whether men's product manufacturers sponsored more blood and gore than other sponsors did.

Gabriel Rossman (2002) had a somewhat different question regarding the mass media. Public concern had grown over the concentration of media in fewer and fewer corporate hands, so Rossman decided to ask the following question: If a newspaper is owned by the same

conglomerate that owns a movie production company, can you trust that newspaper's movie reviews of its parent-company's productions?

You can't, according to Rossman's findings. Because many newspapers rate movies somewhat quantitatively (for example, three stars out of four), he could perform a simple quantitative analysis. For each movie review, he asked two main questions: (1) Was the movie produced by the same company that owned the newspaper? and (2) What rating did the film receive? He found that, indeed, movies produced by the parent company received higher ratings than other movies did. Further, the ratings given to movies by newspapers with the same parent company were higher than the ratings those movies received from other newspapers. This discrepancy, moreover, was strongest in the case of big-budget movies in which the parent company had invested heavily.

As a mode of observation, content analysis requires a thoughtful handling of the "what" that is being communicated. The analysis of data collected in this mode, as in others, addresses the "why" and "with what effect."

Sampling in Content Analysis

In the study of communications, as in the study of people, you often can't observe directly all you would like to explore. In your study of TV violence and sponsorship, for example, I'd advise against attempting to watch everything that's broadcast. It wouldn't be possible, and your brain would probably short-circuit before you came close to discovering that for yourself. Usually, it's appropriate to sample. Let's begin by revisiting the idea of units of analysis. We'll then review some of the sampling techniques that might be applied to such units in content analysis.

Units of Analysis

As I discussed in Chapter 4, determining appropriate units of analysis—the individual units that we make descriptive and explanatory statements about—can be a complicated task. For example, if we wish to compute average family income, the individual family is the unit of analysis. But we'll have to ask individual members of families how much money they make. Thus, individuals will be the units of observation, even though the

individual family remains the unit of analysis. Similarly, we may wish to compare crime rates of different cities in terms of their size, geographic region, racial composition, and other differences. Even though the characteristics of these cities are partly a function of the behaviors and characteristics of their individual residents, the cities would ultimately be the units of analysis.

The complexity of this issue is often more apparent in content analysis than in other research methods, especially when the units of observation differ from the units of analysis. A few examples should clarify this distinction.

Let's suppose we want to find out whether criminal law or civil law makes the most distinctions between men and women. In this instance, individual laws would be both the units of observation and the units of analysis. We might select a sample of a state's criminal and civil laws and then categorize each law by whether or not it makes a distinction between men and women. In this fashion, we could determine whether criminal or civil law distinguishes by gender the most.

Somewhat differently, we might wish to determine whether states that enact laws distinguishing between different racial groups are also more likely than other states to enact laws distinguishing between men and women. Although the examination of this question would also involve the coding of individual acts of legislation, the unit of analysis in this case is the individual state, not the law.

Or, changing topics radically, let's suppose we're interested in representationalism in painting. If we wish to compare the relative popularity of representational and nonrepresentational paintings, the individual paintings will be our units of analysis. If, on the other hand, we wish to discover whether representationalism in painting is more characteristic of wealthy or impoverished painters, of educated or uneducated painters, of capitalist or socialist painters, the individual painters will be our units of analysis.

It's essential that this issue be clear, because sample selection depends largely on what the unit of analysis is. If individual writers are the units of analysis, the sample design should select all or a sample of the writers appropriate to the research question. If books are the units of analysis, we should select a sample of books, regardless of their authors. Bruce Berg (1989: 112–13)

points out that even if you plan to analyze some body of textual materials, the units of analysis might be words, themes, characters, paragraphs, items (such as a book or letter), concepts, semantics, or combinations of these.

I'm not suggesting that sampling should be based solely on the units of analysis. Indeed, we may often subsample—select samples of subcategories—for each individual unit of analysis. Thus, if writers are the units of analysis, we might (1) select a sample of writers from the total population of writers, (2) select a sample of books written by each writer selected, and (3) select portions of each selected book for observation and coding.

Finally, let's look at a trickier example: the study of TV violence and sponsors. What's the unit of analysis for the research question "Are the manufacturers of men's products more likely to sponsor violent shows than other sponsors are?" Is it the TV show? The sponsor? The instance of violence?

In the simplest study design, it would be none of these. Though you might structure your inquiry in various ways, the most straightforward design would be based on the commercial as the unit of analysis. You would use two kinds of observational units: the commercial and the program (the show that gets squeezed in between commercials). You would want to observe both units. You would classify commercials by whether they advertised men's products and the programs by their violence. The program classifications would be transferred to the commercials occurring near them. Figure 11-1 provides an example of the kind of record you might keep.

Notice that in the research design illustrated in Figure 11-1, all the commercials occurring in the same program break are grouped and get the same scores. Also, the number of violent instances recorded as following one commercial break is the same as the number preceding the next break. This simple design allows us to classify each commercial by its sponsorship and the degree of violence associated with it. Thus, for example, the first Grunt Aftershave commercial is coded as being a men's product and as having

Commercial Break	Sponsor	Men's Product?			Number of Instances of Violence	
		Yes	No	?	Before the Commercial Break	After the Commercial Break
1st	Grunt Aftershave	✓			6	4
	Brute Jock Straps	✓			6	4
2nd	Bald-No-More Lotion	✓			4	3
3rd	Grunt Aftershave	✓			3	0
	Snowflake Toothpaste		✓		3	0
	Godliness Cleanser		✓		3	0
4th	Big Thumb Hammers			✓	0	1
5th	Snowflake Toothpaste		✓		1	0
	Big Thumb Hammers			✓	1	0
6th	Buttercup Bras		✓		0	0

FIGURE 11-1

Example of Recording Table for TV Violence

© Cengage Learning®

10 instances of violence associated with it. The Buttercup Bra commercial is coded as not being a men's product and as having no violent instances associated with it.

In the illustration, we have four men's product commercials with an average of 7.5 violent instances each. The four commercials classified as definitely not men's products have an average of 1.75, and the two that might or might not be considered men's products have an average of 1 violent instance each. If this pattern of differences persisted across a much larger number of observations, we'd probably conclude that manufacturers of men's products are more likely to sponsor TV violence than other sponsors are.

The point of this illustration is to demonstrate how units of analysis figure into the data collection and analysis. You need to be clear about your unit of analysis before planning your sampling strategy, but in this case you can't simply sample commercials. Unless you have access to the stations' broadcasting logs, you won't know when the commercials are going to occur. Moreover, you need to observe the programming as well as the commercials. As a result, you must set up a sampling design that will include everything you need in order to observe enough.

In designing the sample, you'd need to establish the universe to be sampled from. In this case, which TV stations will you observe? What will be the period of the study—the number of days? And during which hours of each day will you observe? Then, how many commercials do you want to observe and code for analysis? Watch television for a while and find out how many commercials occur each hour; then you can figure out how many hours of observation you'll need (and can stand).

Now you're ready to design the sample selection. As a practical matter, you wouldn't have to sample among the different stations if you had assistants—each of you could watch a different channel during the same period. But let's suppose you're working alone. Your final sampling frame, from which a sample will be selected and watched, might look something like this:

Jan. 7, Channel 2, 7–9 P.M.
Jan. 7, Channel 4, 7–9 P.M.
Jan. 7, Channel 9, 7–9 P.M.
Jan. 7, Channel 2, 9–11 P.M.

Jan. 7, Channel 4, 9–11 P.M.
Jan. 7, Channel 9, 9–11 P.M.
Jan. 8, Channel 2, 7–9 P.M.
Jan. 8, Channel 4, 7–9 P.M.
Jan. 8, Channel 9, 7–9 P.M.
Jan. 8, Channel 2, 9–11 P.M.
Jan. 8, Channel 4, 9–11 P.M.
Jan. 8, Channel 9, 9–11 P.M.
Jan. 9, Channel 2, 7–9 P.M.
Jan. 9, Channel 4, 7–9 P.M.
etc.

Notice that I've made several decisions for you in the illustration. First, I've assumed that channels 2, 4, and 9 are the ones appropriate to your study. I've assumed that you found the 7–11 P.M. prime-time hours to be the most relevant and that two-hour periods will do the job. I picked January 7 out of the hat for a starting date. In practice, of course, all these decisions should be based on your careful consideration of what would be appropriate to your particular study.

Once you have become clear about your units of analysis and the observations best suited to those units and have created a sampling frame like the one I've illustrated, sampling is simple and straightforward. The alternative procedures available to you are the same ones described in Chapter 7: random, systematic, stratified, and so on.

Sampling Techniques

As we've seen, in the content analysis of written prose, sampling may occur at any or all of several levels, including the contexts relevant to the works. Other forms of communication may also be sampled at any of the conceptual levels appropriate to them.

In content analysis, we could employ any of the conventional sampling techniques discussed in Chapter 7. We might select a random or systematic sample of French and U.S. novelists, of laws passed in the state of Mississippi, or of Shakespearean soliloquies. We might select (with a random start) every 23rd paragraph in Tolstoy's *War and Peace*. Or we might number all of the songs recorded by the Beatles and select a random sample of 25.

Stratified sampling is also appropriate for content analysis. To analyze the editorial policies of U.S. newspapers, for example, we might first group all newspapers by the region of the country or size of the community in which they are published, frequency of publication, or average circulation. We might then select a stratified random or systematic sample of newspapers for analysis. Having done so, we might select a sample of editorials from each selected newspaper, perhaps stratified chronologically.

Cluster sampling is equally appropriate to content analysis. Indeed, if individual editorials are our units of analysis, then the selection of newspapers at the first stage of sampling would be a cluster sample. In an analysis of political speeches, we might begin by selecting a sample of politicians; each politician would represent a cluster of political speeches. The TV-commercial study described previously is another example of cluster sampling.

It should be repeated that sampling need not end when we reach the unit of analysis. If novels are the unit of analysis in a study, we might select a sample of novelists, a subsample of novels written by each selected author, and a subsample of paragraphs within each novel. We would then analyze the content of the paragraphs for the purpose of describing the novels themselves. (If you haven't realized this yet, researchers speak of samples within samples as "subsamples.")

Let's turn now to the coding or classification of the material being observed. Part 4 discusses the manipulation of such classifications to draw descriptive and explanatory conclusions.

Coding in Content Analysis

Content analysis is essentially a coding operation. **Coding** is the process of transforming raw data into a standardized form. In content analysis, communications—oral, written, or other—are coded or classified according to some conceptual framework. Thus, for example, newspaper

---

coding The process whereby raw data are transformed into standardized form suitable for machine processing and analysis.

manifest content In connection with content analysis, the concrete terms contained in a communication, as distinguished from *latent content.*

---

editorials may be coded as liberal or conservative. Radio broadcasts may be coded as propagandistic or not, novels as romantic or not, paintings as representational or not, and political speeches as containing character assassinations or not. Recall that because terms such as these are subject to many interpretations, the researcher must specify definitions clearly.

Coding in content analysis involves the logic of conceptualization and operationalization, which I discussed in Chapter 5. As in other research methods, you must refine your conceptual framework and develop specific methods for observing in relation to that framework.

Manifest and Latent Content

In the earlier discussions of field research, we found that the researcher faces a fundamental choice between depth and specificity of understanding. Often, this represents a choice between validity and reliability, respectively. Typically, field researchers opt for depth, preferring to base their judgments on a broad range of observations and information, even at the risk that another observer might reach a different judgment of the same situation. Survey research—through the use of standardized questionnaires—represents the other extreme: total specificity, even though the specific measures of variables may not be adequately valid reflections of those variables. The content analyst has some choice in this matter, however.

Coding the **manifest content**—the visible, surface content—of a communication is analogous to using a standardized questionnaire. To determine, for example, how erotic certain novels are, you might simply count the number of times the word *love* appears in each novel or the average number of appearances per page. Or, you might use a list of words, such as *love, kiss, hug,* and *caress,* each of which might serve as an indicator of the erotic nature of the novel. This method would have the advantage of ease and reliability in coding and of letting the reader of the research report know precisely how eroticism was measured. It would have a disadvantage, on the other hand, in terms of validity. Surely the phrase *erotic novel* conveys a richer and deeper meaning than the number of times the word *love* is used.

While content analysts in the past have needed to engage in hand counts of terms being

scored, the computer has made this task easier. If you are coding a digital-format document, for example, you could use a search function to jump from one appearance of a term to the next, counting as you went along. However, computer programs such as Wordscores are streamlining this task even further. Let's suppose you would like to code political documents according to political orientations. First, Wordscores can be used to analyze documents of known orientations (e.g., liberal, conservative, etc.), noting what terms tend to be used frequently. Those patterns can then be used to analyze documents you wish to code on the basis of political orientation.

Alternatively, you could code the communication's underlying meaning, or its **latent content**. In the present example, you might read an entire novel or a sample of paragraphs or pages and make an overall assessment of how erotic the novel was. Although your total assessment might very well be influenced by the appearance of words such as *love* and *kiss,* it would not depend fully on their frequency.

Clearly, this second method seems better designed for tapping the underlying meaning of communications, but its advantage comes at a cost to reliability and specificity. Especially if more than one person is coding the novel, somewhat different definitions or standards may be employed. A passage that one coder regards as erotic may not seem erotic to another. Even if you do all of the coding yourself, there is no guarantee that your definitions and standards will remain constant throughout the enterprise. Moreover, the reader of your research report will likely be uncertain about the definitions you've employed. See Figure 11-2 to compare manifest and latent coding.

Carol J. Auster and Claire S. Mansbach (2012) used content analysis to examine differences in the way the Disney Store marketed toys for girls and for boys. First, they noted distinct color differences in the toys: Those for boys tended to have bold colors (e.g., red, black); more of the girls' toys came in pastels, such as pink and purple. Moreover, the toys for boys and for girls tended to reflect traditional stereotypes of masculinity and femininity: Boys had plenty of toys designed for building and killing, whereas girls had toys to prepare them for domestic chores and looking nice.

Manifest Coding of Materials (Objective)

Manifest coding involves the counting of specific elements, such as the word *love*, to determine whether and to what degree the passage should be judged "erotic."
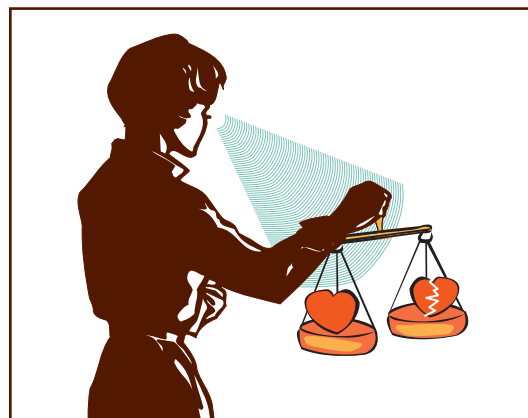
Latent Coding of Materials (Subjective)

Latent coding calls for the researcher to view the entire unit of analysis (a paragraph in this case) and make a subjective assessment regarding whether and to what degree it is "erotic."

FIGURE 11-2

Manifest and Latent Coding

© Cengage Learning®

Conceptualization and the Creation of Code Categories

For all research methods, conceptualization and operationalization typically involve the interaction of theoretical concerns and empirical observations. If, for example, you believe some websites to be liberal and others to be conservative, ask yourself why you think so. Read some

latent content In connection with content analysis, the underlying meaning of communications, as distinguished from their *manifest content.*

content on a few different websites while asking yourself which ones are liberal and which ones are conservative. Things to consider might be: Was the political orientation of a particular editorial most clearly indicated by its manifest content or by its tone? Was your decision based on the use of certain terms (for example, *leftist, fascist,* and so on) or on the support or opposition given to a particular issue or political personality?

Both inductive and deductive methods should be used in this activity. If you're testing theoretical propositions, your theories should suggest empirical indicators of concepts. If you begin with specific empirical observations, you should attempt to derive general principles relating to them and then apply those principles to the other empirical observations.

Bruce Berg (1989: 111) places code development in the context of grounded theory and likens it to solving a puzzle:

> Coding and other fundamental procedures associated with grounded theory development are certainly hard work and must be taken seriously, but just as many people enjoy finishing a complicated jigsaw puzzle, many researchers find great satisfaction in coding and analysis. As researchers . . . begin to see the puzzle pieces come together to form a more complete picture, the process can be downright thrilling.

Throughout this activity, remember that the operational definition of any variable is composed of the attributes included in it. Such attributes, moreover, should be mutually exclusive and exhaustive. A political website, for example, should not be described as both liberal and conservative, though you should probably allow for some to be middle-of-the-road. It may be sufficient for your purposes to code novels as erotic or nonerotic, but you may also want to consider that some could be anti-erotic. Paintings might be classified as representational or not, if that satisfied your research purpose, or you might wish to classify them as impressionistic, abstract, allegorical, and so forth.

Realize further that different levels of measurement can be used in content analysis. You might, for example, use the nominal categories of liberal and conservative for characterizing political websites, or you might wish to use a more refined ordinal ranking, ranging from extremely liberal to extremely conservative. Bear in mind, however, that the level of measurement implicit in your coding methods—nominal, ordinal, interval, or ratio—does not necessarily reflect the nature of your variables. If the word *love* appeared 100 times in Novel A and 50 times in Novel B, you would be justified in saying that the word *love* appeared twice as often in Novel A, but not that Novel A was twice as erotic as Novel B. Similarly, agreeing with twice as many anti-Semitic statements in a questionnaire as someone else does not necessarily make one twice as anti-Semitic as that other person.

Counting and Record Keeping

If you plan to evaluate your content analysis data quantitatively, your coding operation must be amenable to data processing. This means, first, that the end product of your coding must be numerical. If you're counting the frequency of certain words, phrases, or other manifest content, the coding is necessarily numerical. But even if you're coding latent content on the basis of overall judgments, it will be necessary to represent your coding decision numerically: 1 = very liberal, 2 = moderately liberal, 3 = moderately conservative, and so on.

Second, your record keeping must clearly distinguish between units of analysis and units of observation, especially if these two are different. The initial coding, of course, must relate to the units of observation. If novelists are the units of analysis, for example, and you wish to characterize them through a content analysis of their novels, your primary records will represent novels as the units of observation. You may then combine your scoring of individual novels to characterize each novelist, the unit of analysis.

Third, while you're counting, it will normally be important to record the base from which the counting is done. It would probably be useless to know the number of realistic paintings produced by a given artist without knowing the number he or she has painted in total; the painter would be regarded as realistic if a high percentage of paintings were of that genre. Similarly, it would tell us little that the word *love* appeared 87 times in a novel if we did not know about how many words there were in the entire novel. The issue of observational base is most easily resolved if

every observation is coded in terms of one of the attributes making up a variable. Rather than simply counting the number of liberal editorials in a given collection, for example, code each editorial by its political orientation, even if it must be coded "no apparent orientation."

Let's suppose we want to describe and explain the editorial policies of different newspapers. Figure 11-3 presents part of a tally sheet that might result from the coding of newspaper editorials. Note that newspapers are the units of analysis. Each newspaper has been assigned an identification number to facilitate mechanized processing. The second column has a space for the number of editorials coded for each newspaper. This will be an important piece of information, because we want to be able to say, for example, "Of all the editorials, 22 percent were pro–United Nations," not just "There were eight pro–United Nations editorials."

One column in Figure 11-3 is for assigning a subjective overall assessment of each newspaper's editorial policies. (Such assignments might later be compared with the several objective measures.) Other columns provide space for recording numbers of editorials reflecting specific editorial positions. In a real content analysis, there would be spaces for recording other editorial positions plus noneditorial information about each newspaper, such as the region in which it is published, its circulation, and so forth.

The type of content analysis just described is sometimes referred to as *conceptual analysis*, to distinguish it from *relational analysis*. The latter goes beyond observing the frequency of a particular concept in a sample of texts to examining the relationships among concepts. For example, you might look for references to "discrimination" in letters to the editor and also note the kind of discrimination being discussed: racial, religious, gender, and so forth. In fact, you could examine the change in that relationship over time.

Qualitative Data Analysis

Not all content analysis results in counting. Sometimes a qualitative assessment of the materials is most appropriate.

Bruce Berg (1989: 123–25) discusses "negative case testing" as a technique for qualitative hypothesis testing. First, in the grounded theory tradition, you begin with an examination of the data, which may yield a general hypothesis. Let's say that you're examining the leadership of a new community association by reviewing the minutes of meetings to see who made motions that were subsequently passed. Your initial examination of the data suggests that the wealthier

Newspaper ID	Number of editorials evaluated	SUBJECTIVE EVALUATION 1. Very liberal 2. Moderately liberal 3. Middle-of-road 4. Moderately conservative 5. Very conservative	Number of "isolationist" editorials	Number of "pro–United Nations" editorials	Number of "anti–United Nations" editorials
001	37	2	0	8	0
002	26	5	10	0	6
003	44	4	2	1	2
004	22	3	1	2	3
005	30	1	0	6	0

FIGURE 11-3
Sample Tally Sheet (Partial)
© Cengage Learning®

members are the most likely to assume this leadership role.

The second stage in the analysis is to search your data to find all the cases that contradict the initial hypothesis. In this instance, you would look for poorer members who made successful motions and wealthy members who never did. Third, you must review each of the disconfirming cases and either (1) give up the hypothesis or (2) see how it needs to be fine-tuned.

Let's say that in your analysis of disconfirming cases, you notice that each of the unwealthy leaders has a graduate degree, whereas each of the wealthy nonleaders has very little formal education. You may revise your hypothesis to consider both education and wealth as routes to leadership in the association. Perhaps you'll discover some threshold for leadership (a white-collar job, a level of income, and a college degree) beyond which those with the most money, education, or both are the most active leaders.

This process is an example of what Barney Glaser and Anselm Strauss (1967) called *analytic induction.* It is inductive in that it begins primarily with observations, and it is analytic because it goes beyond description to find patterns and relationships among variables.

There are, of course, dangers in this form of analysis, as in all others. The chief risk is misclassifying observations so as to support an emerging hypothesis. For example, you may erroneously conclude that a nonleader didn't graduate from college or you may decide that the job of factory foreman is "close enough" to being white-collar. Berg (1989: 124) offers techniques for avoiding these errors:

1. If there are sufficient cases, select some at random from each category in order to avoid merely picking those that best support the hypothesis.

2. Give at least three examples in support of every assertion you make about the data.

3. Have your analytic interpretations carefully reviewed by others uninvolved in the research project to see whether they agree.

4. Report whatever inconsistencies you do discover—any cases that simply do not fit your hypotheses. Realize that few social patterns are 100 percent consistent, so you may have discovered something important even if

it doesn't apply to absolutely all of social life. However, you should be honest with your readers in that regard.

There are computer programs now available for content analysis. For example, you can try out MAXQDA online. Also, T-LAB provides for some interesting qualitative analyses, such as mapping word associations in a political speech. Some of the programs appropriate for content analysis are discussed in Chapter 13 in connection with other kinds of qualitative data analysis.

Illustrations of Content Analysis

Several studies have indicated that historically women have been stereotyped on television. R. Stephen Craig (1992) took this line of inquiry one step further to examine the portrayal of both men and women during different periods of television programming.

To study gender stereotyping in television commercials, Craig selected a sample of 2,209 network commercials during several periods between January 6 and 14, 1990.

> The weekday day part (in this sample, Monday–Friday, 2–4 P.M.) consisted exclusively of soap operas and was chosen for its high percentage of women viewers. The weekend day part (two consecutive Saturday and Sunday afternoons during sports telecasts) was selected for its high percentage of men viewers. Evening "prime time" (Monday–Friday, 9–11 P.M.) was chosen as a basis for comparison with past studies and the other day parts.
>
> *(1992: 199)*

Each of the commercials was coded in several ways. "Characters" were coded as

All male adults
All female adults
All adults, mixed gender
Male adults with children or teens (no women)
Female adults with children or teens (no men)
Mixture of ages and genders

In addition, Craig's coders noted which character was on the screen longest during the commercial—the "primary visual character"—as well as the roles played by the characters (such

TABLE 11-1

Percent of Adult Primary Visual Characters by Gender Appearing in Commercials at Different Times

	Weekend	Daytime	Evening
Adult male	40	52	80
Adult female	60	48	20

Source: R. Stephen Craig. 1992. "The Effect of Television Day Part on Gender Portrayals in Television Commercials: A Content Analysis," *Sex Roles* 26 (5/6): 204.

as spouse, celebrity, parent), the type of product advertised (such as body product, alcohol), the setting (such as kitchen, school, business), and the voice-over narrator.

Table 11-1 indicates the differences in the times when men and women appeared in commercials. Women appeared most during the daytime (with its soap operas), men predominated during the weekend commercials (with its sports programming), and men and women were equally represented during evening prime time.

Craig found other differences in the ways men and women were portrayed.

> Further analysis indicated that male primary characters were proportionately more likely than females to be portrayed as celebrities and professionals in every day part, while women were proportionately more likely to be portrayed as interviewer/demonstrators, parent/spouses, or sex object/models in every day part. . . . Women were proportionately more likely to appear as sex object/models during the weekend than during the day.
>
> *(1992: 204)*

The research also showed that different products were advertised during different time periods. As you might imagine, almost all the daytime commercials dealt with body, food, or home products. These products accounted for only one in three on the weekends. Instead, weekend commercials stressed automotive products (29 percent), business products or services (27 percent), or alcohol (10 percent). There were virtually no alcohol ads during evenings and daytime.

As you might suspect, women were most likely to be portrayed in home settings, men most likely to be shown away from home. Other findings dealt with the different roles played by men and women.

> The women who appeared in weekend ads were almost never portrayed without men and seldom as the commercial's primary character. They were generally seen in roles subservient to men (e.g., hotel receptionist, secretary, or stewardess), or as sex objects or models in which their only function seemed to be to lend an aspect of eroticism to the ad.
>
> *(Craig 1992: 208)*

Although some of Craig's findings may seem unsurprising, remember that "common knowledge" does not always correspond with reality. It's always worthwhile to check out widely held assumptions. And even when we think we know about a given situation, it's often useful to know specific details such as those provided by a content analysis like this one.

In another content analysis that drew on popular culture for content, Charis Kubrin (2005) chose a primarily qualitative approach. Kubrin was interested in the themes put forth in rap music, particularly in gangsta rap, and the relationship of those themes to neighborhood culture and the "street code."

> In response to societal and neighborhood conditions, black youth in disadvantaged communities have created a substitute social order governed by their own code—a street code—and rituals of authenticity. . . . This social order reflects the subcultural locus of interests that emerges from pervasive race and class inequality and the social isolation of poor black communities.
>
> *(2005: 439)*

She began her study by identifying all the platinum rap albums released between 1992 and 2000: 130 albums containing a total of 1,922 songs. She then drew a simple random sample of one-third of the songs (632) and set about the task of listening to each. She did this twice with each song.

> First, I listened to a song in its entirety while reading the printed lyrics to determine what the song was about. Second, I listened to the song again and coded each line to determine whether the street code elements described earlier were present: (1) respect, (2) willingness to fight or use violence, (3) material wealth, (4) violent retaliation, (5) objectification of women, and (6) nihilism.
>
> *(2005: 443)*

Kubrin was particularly interested in the theme of nihilism—the rejection of traditional moral principles and a fundamental skepticism about the meaning of life. She was interested in how that theme was portrayed in gangsta rap and how it fit into the street code.

Though she began with a sample of 632 songs, she found that no new themes appeared to be showing up after about 350 songs had been analyzed. To be safe, she coded another 50 songs and found no new themes, completing her coding process at that point.

Kubrin notes that rap music is typically regarded as antisocial and resistant to organized society, but her in-depth analysis of lyrics suggests something different:

> Rap music does not exist in a cultural vacuum; rather it expresses the cultural crossing, mixing, and engagement of black youth culture with the values, attitudes and concerns of the white majority. Many of the violent (and patriarchal, materialistic, sexist, etc.) ways of thinking that are glorified in gangsta rap are a reflection of the prevailing values created and sustained in the larger society.
>
> *(2005: 454)*

She traces the implications of this for understanding street life as well as for the likely success of various crime-control strategies.

Strengths and Weaknesses of Content Analysis

Probably the greatest advantage of content analysis is its economy in terms of both time and money. A college student might undertake a content analysis, whereas undertaking a survey, for example, might not be feasible. There is no requirement for a large research staff; no special equipment is needed. As long as you have access to the material to be coded, you can undertake content analysis.

Content analysis also has the advantage of allowing the correction of errors. If you discover you've botched up a survey or an experiment, you may be forced to repeat the whole research project with all its attendant costs in time and money. If you botch up your field research, it may be impossible to redo the project; the event under study may no longer exist. In content

analysis, it's usually easier to repeat a portion of the study than it is in other research methods. You might be required, moreover, to recode only a portion of your data rather than all of it.

A third advantage of content analysis is that it permits the study of processes occurring over a long time. You might focus on the imagery of Irish Americans conveyed in U.S. novels written between 1850 and 1860, for example, or you might examine how such imagery has changed from 1850 to the present.

Finally, content analysis has the advantage of all unobtrusive measures, namely, that the content analyst seldom has any effect on the subject being studied. Because the novels have already been written, the paintings already painted, the speeches already presented, content analyses can have no effect on them.

Content analysis has disadvantages as well. For one thing, it's limited to the examination of recorded communications. Such communications may be oral, written, or graphic, but they must be recorded in some fashion to permit analysis.

As we've seen, content analysis has both advantages and disadvantages in terms of validity and reliability. Problems of validity are likely unless you happen to be studying communication processes per se.

On the other side of the ledger, the concreteness of materials studied in content analysis strengthens the likelihood of reliability. You can always code your data and then recode the original documents from scratch. And you can repeat the process as many times as you want. In field research, by contrast, there's no way to return to the original events that were observed, recorded, and categorized.

Let's move from content analysis now and turn to a related research method: the analysis of existing data. Although numbers rather than communications are analyzed in this case, I think you'll see the similarity to content analysis.

Analyzing Existing Statistics

Frequently you can or must undertake social science inquiry through the use of official or quasi-official statistics. This differs from secondary analysis, in which you obtain a copy of someone else's data and undertake your own

statistical analysis. In this section, we're going to look at ways of using the data analyses that others have already done.

This method is particularly significant because existing statistics should always be considered as at least a supplemental source of data. If you were planning a survey of political attitudes, for example, you would do well to examine and present your findings within a context of voting patterns, rates of voter turn-out, or similar statistics relevant to your research interest. Or, if you were doing evaluation research on an experimental morale-building program on an assembly line, then statistics on absenteeism, sick leave, and so on would probably be interesting and revealing in connection with the data from your own research. Existing statistics, then, can often provide a historical or conceptual context within which to locate your original research.

Existing statistics can also provide the main data for a social science inquiry. An excellent example is the classic study mentioned at the beginning of this chapter, Emile Durkheim's *Suicide* ([1897] 1951). Let's take a closer look at Durkheim's work before considering some of the special problems this method presents.

Durkheim's Study of Suicide

Why do people kill themselves? Undoubtedly every suicide case has a unique history and ex-planation, yet all such cases could no doubt be grouped according to certain common causes: financial failure, trouble in love, disgrace, and other kinds of personal problems. The French so-ciologist Emile Durkheim had a slightly different question in mind when he addressed the matter of suicide, however. He wanted to discover the environmental conditions that encouraged or discouraged it, especially social conditions.

The more Durkheim examined the avail-able records, the more patterns of differences became apparent to him. One of the first things to attract his attention was the relative stability of suicide rates. Looking at several countries, he found suicide rates to be about the same year after year. He also discovered that a dispropor-tionate number of suicides occurred in summer, leading him to hypothesize that temperature might have something to do with it. If this were

the case, suicide rates should be higher in the southern European countries than in the tem-perate ones. However, Durkheim discovered that the highest rates were found in countries in the central latitudes, so temperature could not be the answer.

He explored the role of age (35 was the most common suicide age), gender (men outnum-bered women around four to one), and numer-ous other factors. Eventually, a general pattern emerged from different sources.

In terms of the stability of suicide rates over time, for instance, Durkheim found that the pat-tern was not totally stable. There were spurts in the rates during times of political turmoil, which occurred in several European countries around 1848. This observation led him to hypothesize that suicide might have something to do with "breaches in social equilibrium." Put differently, social stability and integration seemed to be a protection against suicide.

This general hypothesis was substantiated and specified through Durkheim's analysis of a different set of data. The different countries of Europe had radically different suicide rates. The rate in Saxony, for example, was about ten times that of Italy, and the relative ranking of vari-ous countries persisted over time. As Durkheim considered other differences among the various countries, he eventually noticed a striking pat-tern: Predominantly Protestant countries had consistently higher suicide rates than Catholic ones did. The predominantly Protestant countries had 190 suicides per million population; mixed Protestant-Catholic countries, 96; and predomi-nantly Catholic countries, 58 (Durkheim [1897] 1951: 152).

Although suicide rates thus seemed to be related to religion, Durkheim reasoned that some other factor, such as level of economic and cultural development, might explain the observed differences among countries. If religion had a genuine effect on suicide, then the religious difference would have to be found *within* given countries as well. To test this idea, Durkheim first noted that the German state of Bavaria had both the most Catholics and the lowest suicide rates in that country, whereas heavily Protestant Prussia had a much higher suicide rate. Not content to stop there, however, Durkheim examined the provinces composing each of those states.

Table 11-2 shows what he found. As you can see, in both Bavaria and Prussia, provinces with the highest proportion of Protestants also had the highest suicide rates. Increasingly, Durkheim became confident that religion played a significant role in the matter of suicide.

Returning eventually to a more general theoretical level, Durkheim combined the religious findings with the earlier observation about increased suicide rates during times of political turmoil. As we've seen, Durkheim suggested that many suicides are a product of *anomie,* that is, "normlessness," or a general sense of social instability and disintegration. During times of political strife, people may feel that the old ways of society are collapsing. They become demoralized and depressed, and suicide is one answer to the severe discomfort. Seen from the other direction, social integration and solidarity—reflected in personal feelings of being part of a coherent, enduring social whole—offer protection against depression and suicide. That was where the religious difference fit in. Catholicism, as a far more structured and integrated religious system, gave people a greater sense of coherence and stability than did the more loosely structured Protestantism.

From these theories, Durkheim created the concept of *anomic suicide.* More importantly, as you may know, he added the concept of *anomie* to the lexicon of the social sciences.

This account of Durkheim's classic study is greatly simplified, of course. Anyone studying social research would profit from studying the original. For our purposes, Durkheim's approach provides a good illustration of the possibilities for research contained in the masses of data regularly gathered and reported by government agencies and other organizations.

In a more recent examination of suicide rates, Steven Barkan and his colleagues (2013) try to explain the relatively higher rates of suicide in the American West. Reminiscent of Durkheim's conclusion regarding social solidarity, the researchers found that residential stability was a strong force for lowering suicide rates.

The Consequences of Globalization

The notion of "globalization" has become increasingly controversial in the United States and around the world, with reactions ranging from

TABLE 11-2

Suicide Rates in Various German Provinces, Arranged in Terms of Religious Affiliation

Religious Character of Province	Suicides per Million Inhabitants
*Bavarian Provinces (1867–1875)**	
Less than 50% Catholic	
Rhenish Palatinate	167
Central Franconia	207
Upper Franconia	204
Average	**192**
50% to 90% Catholic	
Lower Franconia	157
Swabia	118
Average	**135**
More than 90% Catholic	
Upper Palatinate	64
Upper Bavaria	114
Lower Bavaria	19
Average	**75**
Prussian Provinces (1883–1890)	
More than 90% Protestant	
Saxony	309.4
Schleswig	312.9
Pomerania	171.5
Average	**264.6**
68% to 89% Protestant	
Hanover	212.3
Hesse	200.3
Brandenburg and Berlin	296.3
East Prussia	171.3
Average	**220.0**
40% to 50% Protestant	
West Prussia	123.9
Silesia	260.2
Westphalia	107.5
Average	**163.6**
28% to 32% Protestant	
Posen	96.4
Rhineland	100.3
Hohenzollern	90.1
Average	**95.6**

*Note: The population below 15 years of age has been omitted.

Source: Adapted from Emile Durkheim, *Suicide* (Glencoe, IL: Free Press, [1897] 1951), 153.

scholarly debates to violent confrontations in the streets. One point of view sees the spread of U.S.-style capitalism to developing countries as economic salvation for those countries. A very different point of view sees globalization as essentially neocolonial exploitation, in which multinational conglomerates exploit the resources and people of poor countries. And, of course, there are numerous variations on these contradictory views.

Jeffrey Kentor (2001) wanted to bring data to bear on the question of how globalization affects the developing countries that host the process. To that end, he used data available from the World Bank's "World Development Indicators." Noting past variations in the way globalization was measured, Kentor used the amount of foreign investment in a country's economy as a percentage of that country's whole economy. He reasoned that dependence on foreign investments was more important than the amount of the investment.

In his analysis of 88 countries with a per capita gross domestic product (the total goods and services produced in a country) of less than $10,000, Kentor found that dependence on foreign investment tended to increase income inequality among the citizens of a country. The greater the degree of dependence, the greater the income inequality. Kentor reasoned that globalization produced well-paid elites who, by working with the foreign corporations, maintained a status well above that of the average citizen. But because the profits derived from the foreign investments tended to be returned to the investors' countries rather than enriching the poor countries, the great majority of the population in the latter reaped little or no economic benefit.

Income inequality, in turn, was found to increase birth rates and, hence, population growth, in a process too complex to summarize here. Population growth, of course, brings a wide range of problems to countries already too poor to provide for the current basic needs of their people.

This research example, along with our brief look at Durkheim's studies, should broaden your understanding of the kinds of social phenomena that we can study through data already collected and compiled by others.

Units of Analysis

The unit of analysis involved in the analysis of existing statistics is often not the individual. Durkheim, for example, was required to work with political–geographic units: countries, regions, states, and cities. The same situation would probably appear if you were to undertake a study of crime rates, accident rates, or disease. By their nature, most existing statistics are aggregated: They describe groups.

The aggregate nature of existing statistics can present a problem, though not an insurmountable one. As we saw, for example, Durkheim wanted to determine whether Protestants or Catholics were more likely to commit suicide. The difficulty was that none of the records available to him indicated the religion of those people who committed suicide. Ultimately, then, it was not possible for him to say whether Protestants committed suicide more often than Catholics did, though he inferred as much. Because Protestant countries, regions, and states had higher suicide rates than did Catholic countries, regions, and states, he drew the obvious conclusion.

There's danger in drawing this kind of conclusion, however. It's always possible that patterns of behavior at a group level do not reflect corresponding patterns on an individual level. Such errors are due to an ecological fallacy, which was discussed in Chapter 4. In the case of Durkheim's study, it was altogether possible, for example, that it was Catholics who committed suicide in the predominantly Protestant areas. Perhaps Catholics in predominantly Protestant areas were so badly persecuted that they were led into despair and suicide. In that case it would be possible for Protestant countries to have high suicide rates without any Protestants committing suicide.

Durkheim avoided the danger of the ecological fallacy in two ways. First, his general conclusions were based as much on rigorous theoretical deductions as on the empirical facts. The correspondence between theory and fact made a counter explanation, such as the one I just made up, less likely. Second, by extensively retesting his conclusions in a variety of ways, Durkheim further strengthened the likelihood that they were correct. Suicide rates were higher in Protestant countries than in Catholic ones; higher in Protestant regions of Catholic countries than

in Catholic regions of Protestant countries; and so forth. The replication of findings added to the weight of evidence in support of his conclusions.

Problems of Validity

Whenever we base research on an analysis of data that already exist, we're obviously limited to what already exists. Often, the existing data do not cover exactly what we're interested in, and our measurements may not be altogether valid representations of the variables and concepts we want to make conclusions about.

Two characteristics of science are used to handle the problem of validity in analysis of existing statistics: *logical reasoning* and *replication.* Durkheim's strategy provides an example of logical reasoning. Although he could not determine the religion of people who committed suicide, he reasoned that most of the suicides in a predominantly Protestant region would be Protestants.

Replication can be a general solution to problems of validity in social research. Recall the earlier discussion of the interchangeability of indicators (Chapter 5). Crying in sad movies isn't necessarily a valid measure of compassion; nor is putting little birds back in their nests nor giving money to charity. None of these things, taken alone, would prove that one group (women, say) was more compassionate than another (men). But if women appeared more compassionate than men by all these measures, that would create a weight of evidence in support of the conclusion. In the analysis of existing statistics, a little ingenuity and reasoning can usually turn up several independent tests of a given hypothesis. If all the tests seem to confirm the hypothesis, then the weight of evidence supports the validity of the measure.

Problems of Reliability

The analysis of existing statistics depends heavily on the quality of the statistics themselves: Do they accurately report what they claim to report? This can be a substantial problem sometimes, because the weighty tables of government statistics, for example, are sometimes grossly inaccurate.

Consider research into crime. Because a great deal of this research depends on official crime statistics, this body of data has come under critical evaluation. The results have not been encouraging. As an illustration, suppose you were interested in tracing long-term trends in marijuana use in the United States. Official statistics on the numbers of people arrested for selling or possessing marijuana would seem to be a reasonable measure of use, right? Not necessarily.

To begin, you face a hefty problem of validity. Before the passage of the Marihuana Tax Act in 1937, "grass" was legal in the United States, so arrest records would not give you a valid measure of use. But even if you limited your inquiry to the times after 1937, you would still have problems of reliability that stem from the nature of law enforcement and crime recording, not to mention those states, such as Washington and Colorado, who have legalized all uses of it.

Law enforcement, for example, is subject to various pressures. A public outcry against marijuana, led perhaps by a vocal citizens' group, often results in a police crackdown on drug trafficking—especially during an election or budget year. A sensational story in the press can have a similar effect. In addition, the volume of other business facing the police can affect marijuana arrests.

In tracing the pattern of drug arrests in Chicago between 1942 and 1970, Lois DeFleur (1975) demonstrates that the official records present a far less accurate history of drug use than of police practices and political pressure on police. On a different level of analysis, Donald Black (1970) and others have analyzed the factors influencing whether an offender is actually arrested by police or let off with a warning. Ultimately, official crime statistics are influenced by whether specific offenders are well or poorly dressed, whether they are polite or abusive to police officers, and so forth. When we consider unreported crimes, sometimes estimated to be as much as ten times the number of crimes known to police, the reliability of crime statistics gets even shakier.

These comments concern crime statistics at a local level. Often it's useful to analyze national crime statistics, such as those reported in the FBI's annual *Uniform Crime Reports.* Additional problems are introduced at the national level. For example, different local jurisdictions define crimes differently. Also, participation in the FBI program is voluntary, so the data are incomplete.

Finally, the process of record keeping affects the data available to researchers. Whenever a law-enforcement unit improves its record-keeping system—computerizes it, for example—the apparent crime rates increase dramatically. This can happen even if the number of crimes committed, reported, and investigated does not increase.

Researchers' first protection against the problems of reliability in the analysis of existing statistics is knowing that the problem may exist. Investigating the nature of the data collection and tabulation may enable you to assess the nature and degree of unreliability so that you can judge its potential impact on your research interest. If you also use logical reasoning and replication, you can usually cope with the problem.

Sources of Existing Statistics

It would take a whole book just to list the sources of data available for analysis. In this section, I want to mention a few sources and point you in the direction of finding others relevant to your research interest.

Undoubtedly, one of the more important resources of U.S.-focused data is the annual *Statistical Abstract of the United States*, published by the Department of Commerce. This wealth of information includes statistics on the individual states and (less extensively) cities, as well as on the nation as a whole. Where else can you find the number of work stoppages in the country year by year, the residential property taxes of major cities, the number of water-pollution discharges reported around the country, the number of business proprietorships in the nation, and hundreds of other such handy bits of information? Best of all, you can download the *Statistical Abstract* from the web for free (your tax dollars at work for you).

While you are probably most familiar with the U.S. Census in terms of its role in the decennial enumeration of the whole population, as prescribed by the Constitution, the Census Bureau conducts numerous other studies. The American Community Survey is another useful source, employing more-frequent sample surveys of nation. You should be able to learn about where you live, although the extent and accuracy of the data will depend on the size of

your community. You can also use the online program, Census Explorer, to examine the American Community Survey data.

Suppose you were interested in the issue of income discrimination by gender. You could examine this rather easily through the *Statistical Abstract* data. The following table, for example, provides a look at gender, education, and income (adapted from U.S. Bureau of the Census 2012: Table 703, p. 459). As you can see, women have still not reached a parity with men, even when they have the same level of education.

Average Earnings of Year-Round, Full-Time Workers, 2009

	Men	Women	Ratio of Women/Men Earnings
All workers	$62,445	$44,857	0.72
Less than 9th grade	$26,604	$19,588	0.74
9th–12th grades	$33,194	$23,478	0.71
H.S. graduates	$43,140	$32,227	0.75
Some college	$52,580	$36,553	0.70
Associate degree	$55,631	$42,307	0.76
Bachelor's or more	$92,815	$62,198	0.67

Source: U.S. Bureau of the Census. 2012. *Statistical Abstract of the United States.* Table 703, p. 459. Washington, DC: U.S. Government Printing Office. You can also access this table online at http://www.census.gov/compendia/statab/2012/tables/12s0703.pdf.

As we've seen before, a graphic presentation can sometimes communicate data more easily than tables of numbers. You could enter the above incomes into a spreadsheet program and have it create a graphic display as shown Figure 11-4.

These data point to a persistent difference between the incomes of men and women, even when both groups have achieved the same levels of education. Other variables could explain the differences, however; we'll return to this issue in Chapter 15.

World statistics are available through the United Nations. Its *Demographic Yearbook* presents annual vital statistics (births, deaths, and other data relevant to population) for the individual nations of the world. Other publications report a variety of other kinds of data. Again, utilizing the resources at your library and on the web may be the best introduction to what's available.
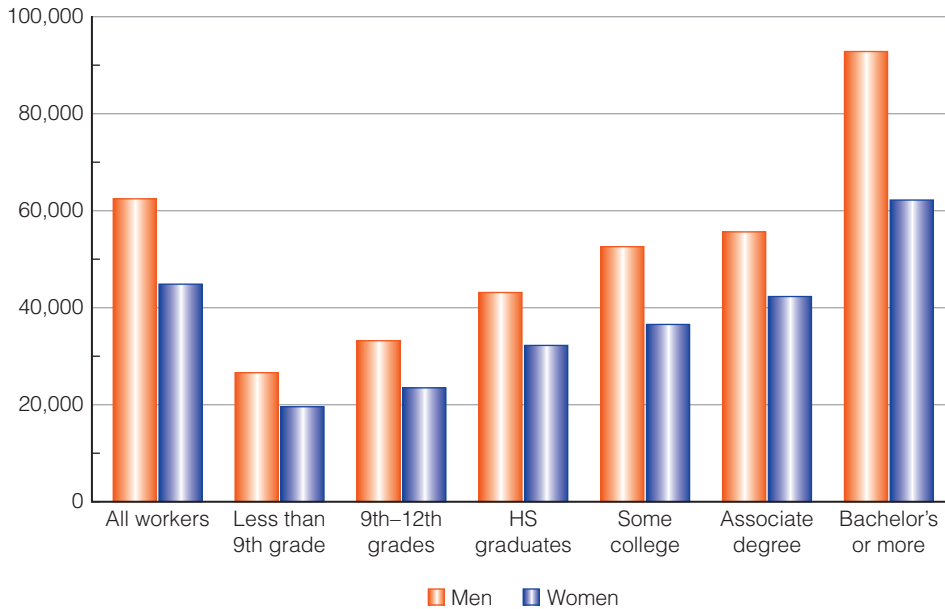
FIGURE 11-4

Graphic Display of Gender, Education, and Income

© Cengage Learning®

The amount of data provided by nongovernment agencies is as staggering as the amount your taxes buy. Chambers of commerce often publish data reports on businesses, as do private consumer groups. Common Cause covers politics and government. The Gallup Organization publishes reference volumes on public opinion as tapped by Gallup Polls since 1935.

Organizations such as the Population Reference Bureau publish a variety of demographic data, U.S. and international, that a secondary analyst could use. Their *World Population Data Sheet* and *Population Bulletin* are resources heavily used by social scientists. Social indicator data can be found in the journal *SINET: A Quarterly Review of Social Reports and Research on Social Indicators, Social Trends, and the Quality of Life*.

A new guide to Population Action International's mapping website shows how climate change and population dynamics will change the world over time. High rates of population growth and climate-change consequences overlap in many countries. Interactive maps illustrate how climate-change impacts, demographic trends, and the need for contraception are likely to affect countries' abilities to adapt to the effects of climate change.

The maps identify 33 population and climate-change hotspots—countries that are experiencing rapid population growth, low resilience to climate change, and high projected declines in agricultural production. Many hotspots are currently experiencing water stress or scarcity, a condition that will worsen with continued rapid population growth. And in many countries, a high proportion of women lack access to reproductive health services and contraceptives. Investments in family-planning programs in these hotspots could improve health and well-being, slow population growth, and reduce vulnerability to climate-change impacts. The newly updated interactive mapping website can be viewed at www.populationaction.org/climatemap.

The sources I've listed represent only a tiny fraction of the thousands that are available. With so much data already collected, the lack of funds to support expensive data collection is no reason for not doing good and useful social research. Moreover, as we've seen, this research method need not be limited to tables of numbers. There are graphic resources available as well, such as the *Social Explorer.* A wide range of data about the United States can be represented on a map of congressional districts or census tracts. You can examine aspects of population, religion, economy, and many other variables. For example, you can easily find the geographic

concentrations of unmarried partners: male/female, male/male, and female/female.

You can do similar kinds of map-based examinations through the Census Bureau by clicking on "Maps" at their website. Once you've displayed a variable such as multiracial marriages state-by-state, you can click on a particular state and get a detailed graph of the racial marriages in that state.

Let's move now from an inherently quantitative method to one that is typically qualitative: comparative and historical research.

Comparative and Historical Research

Comparative and historical research differs substantially from the methods discussed so far, though it overlaps somewhat with field research, content analysis, and the analysis of existing statistics. It involves the use of historical methods by sociologists, political scientists, and other social scientists to examine societies (or other social units) over time and in comparison with one another.

The discussion of longitudinal research designs in Chapter 4 notwithstanding, our examination of research methods so far has focused primarily on studies anchored in one point in time and in one locale, whether a small group or a nation. Although accurately portraying the main thrust of contemporary social science research, this focus conceals the fact that social scientists are also interested in tracing the development of social forms over time and comparing those developmental processes across cultures. James Mahoney and Dietrich Rueschemeyer (2003: 4) suggest that current comparative and historical researchers "focus on a wide range of topics, but they are united by a commitment to providing historically grounded explanations of large-scale and substantively important outcomes." Thus, you find comparative and historical studies dealing with the topics social class, capitalism, religion, revolution, and the like.

After describing some major instances of comparative and historical research, past and present, this section discusses some of the key elements of this method.

Examples of Comparative and Historical Research

Auguste Comte, who coined the term *sociologie,* saw that new discipline as the final stage in a historical development of ideas. With his broadest brush, he painted an evolutionary picture that took humans from a reliance on religion to metaphysics to science. With a finer brush, he portrayed science as evolving from the development of biology and the other natural sciences to the development of psychology and, finally, to the development of scientific sociology.

A great many later social scientists have also turned their attention to broad historical processes. Several have examined the historical progression of social forms from the simple to the complex, from rural-agrarian to urban-industrial societies. The U.S. anthropologist Lewis Morgan, for example, saw a progression from "savagery" to "barbarism" to "civilization" (1870). Seventy years later, Robert Redfield, another anthropologist, wrote of a shift from "folk society" to "urban society" (1941). Emile Durkheim saw social evolution largely as a process of ever-greater division of labor ([1893] 1964). In a more specific analysis, Karl Marx examined economic systems progressing historically from primitive to feudal to capitalistic forms ([1867] 1967). All history, he wrote in this context, was a history of class struggle—the "haves" struggling to maintain their advantages and the "have-nots" struggling for a better lot in life. Looking beyond capitalism, Marx saw the development of socialism and finally communism.

Not all historical studies in the social sciences have had this evolutionary flavor, however. Some social science readings of the historical record, in fact, point to grand cycles rather than to linear progressions. No scholar better represents this view than Pitirim A. Sorokin. A participant in the Russian Revolution of 1917, Sorokin served as secretary to Prime Minister Kerensky. Both Kerensky and Sorokin fell from favor, however, and Sorokin began his second career—as a U.S. sociologist.

comparative and historical research The examination of societies (or other social units) over time and in comparison with one another.

Whereas Comte read history as a progression from religion to science, Sorokin (1937–1940) suggested that societies alternate cyclically between two points of view, which he called "ideational" and "sensate." Sorokin's sensate point of view defined reality in terms of sense experiences. The ideational, by contrast, placed a greater emphasis on spiritual and religious factors. Sorokin's reading of the historical record further indicated that the passage between the ideational and sensate was through a third point of view, which he called the "idealistic." This third view combined elements of the sensate and ideational in an integrated, rational view of the world.

These examples indicate some of the topics comparative and historical researchers have examined. To get a better sense of what comparative and historical research entails, let's look at a few examples in somewhat more detail.

Weber and the Role of Ideas

In his analysis of economic history, Karl Marx put forward a view of economic determinism. That is, he postulated that economic factors determined the nature of all other aspects of society. For example, Marx's analysis showed that a function of European churches was to justify and support the capitalist status quo— religion was a tool of the powerful in maintaining their dominance over the powerless. "Religion is the sigh of the oppressed creature," Marx wrote in a famous passage, "the sentiment of a heartless world, and the soul of soulless conditions. It is the opium of the people" (Bottomore and Rubel [1843] 1956: 27).

Max Weber, a German sociologist, disagreed. Without denying that economic factors could and did affect other aspects of society, Weber argued that economic determinism did not explain everything. Indeed, Weber said, economic forms could come from noneconomic ideas. In his research in the sociology of religion, Weber examined the extent to which religious institutions were the source of social behavior rather than mere reflections of economic conditions. His most noted statement of this side of the issue is found in *The Protestant Ethic and the Spirit of Capitalism* ([1905] 1958). Here's a brief overview of Weber's thesis.

John Calvin (1509–1564), a French theologian, was an important figure in the Protestant reformation of Christianity. Calvin taught that the ultimate salvation or damnation of every individual had already been decided by God; this idea is called *predestination*. Calvin also suggested that God communicated his decisions to people by making them either successful or unsuccessful during their earthly existence. God gave each person an earthly "calling"—an occupation or profession— and manifested their success or failure through that medium. Ironically, this point of view led Calvin's followers to seek proof of their coming salvation by working hard, saving their money, and generally striving for economic success.

In Weber's analysis, Calvinism provided an important stimulus for the development of capitalism. Rather than "wasting" their money on worldly comforts, the Calvinists reinvested it in their economic enterprises, thus providing the capital necessary for the development of capitalism. In arriving at this interpretation of the origins of capitalism, Weber researched the official doctrines of the early Protestant churches, studied the preaching of Calvin and other church leaders, and examined other relevant historical documents.

In three other studies, Weber conducted detailed historical analyses of Judaism ([1934] 1952) and the religions of China ([1934] 1951) and India ([1934] 1958). Among other things, Weber wanted to know why capitalism had not developed in the ancient societies of China, India, and Israel. In none of the three religions did he find any teaching that would have supported the accumulation and reinvestment of capital—strengthening his conclusion about the role of Protestantism in that regard.

Fair Trade

If you buy coffee at a grocery store or coffeehouse, you may have noticed that some of the packages are labeled "fair trade." As you might know, the fair trade certification reflects an international, social/ecological/economic movement formed to support farmers and laborers in developing countries. The fair trade movement seeks equity in international trade, and aims to ensure that these workers receive a higher price for the products they grow and export. In a free-market economy, it is common that growers of products like coffee, chocolate, and bananas actually receive very little of the money that you,

a consumer in a developed country, might pay for it. In practice, fair trade reflects economic reorganization. It may include local farmer co-ops working with international nonprofit organizations, such as the Institute for Agriculture and Trade Policy, to cut out the "middlemen" and thus deliver more money and price stability to those doing the work. Fair trade practices are also focused on improving environmental standards and sustainability practices.

Daniel Jaffee (2007) came in contact with that movement in 2003 while attending a meeting of the World Trade Organization in Mexico. A group of the delegates staged a demonstration on behalf of fair trade and walked out of the WTO meeting to move into a smaller conference of their own. Jaffee followed them and began his extended study of fair trade economics.

> Over two years, I lived, worked, and talked with these farmers, as well as with their neighbors who know a very different coffee market—the conventional market represented by local coyotes, middlemen who often pay them less than it costs to produce their coffee in the first place."
>
> *(2007: xiv)*

Jaffee's research involved participant observation, as his description indicates, but also the collection and analysis of quantitative data about production, prices, income, and the like. In part, he was interested in placing the new movement within the larger context of world coffee production and marketing. (Fair trade presents roughly 1 percent of the total.)

He was also interested in the evolution of the movement over time, as fair trade became better known and more popular. He examined the development of the organizations involved and looked at the adjustments required when large distributors such as Starbucks began offering fair trade coffee as an option for its customers. Whereas we have seen that some research methods offer a snapshot of social life at one point in time, Jaffee's analysis offers a motion picture of an ongoing social process.

Here are a few briefer examples to illustrate some of the topics interesting to comparative and historical scholars today.

- *The Rise of Christianity:* Rodney Stark (1997) lays out his research question in the book's

subtitle: *How the Obscure, Marginal Jesus Movement Became the Dominant Religious Force in the Western World in a Few Centuries.* For many people, the answer to this puzzle is a matter of faith in the miraculous destiny of Christianity. Without debunking Christian faith, Stark looks for a scientific explanation, undertaking an analysis of existing historical records that sketch out the population growth of Christianity during its early centuries. He notes, among other things, that the early growth rate of Christianity, rather than being unaccountably rapid, was very similar to the contemporary growth of Mormonism. He then goes on to examine elements in early Christian practice that gave it growth advantages over the predominant paganism of the Roman Empire. For example, the early Christian churches were friendlier to women than paganism was, and much of the early growth occurred among women—who often converted their husbands later on. And in an era of deadly plagues, the early Christians were more willing to care for stricken friends and family members, which not only enhanced the survival of Christians but also made it a more attractive conversion prospect. At every turn in the analysis, Stark makes rough calculations of the demographic impact of cultural factors. This study is an illustration of how social research methods can shed light on nonscientific realms such as faith and religion.

- *Policing World Society:* Mathieu Deflem (2002) set out to learn how contemporary systems of international cooperation among police agencies came about. All of us have heard movie and TV references to the international police organization, Interpol. Deflem went back to the middle of the nineteenth century and traced its development through World War II. In part, his analysis examines the strains between the bureaucratic integration of police agencies in their home governments and the need for independence from those governments.

- *Organizing America:* Charles Perrow (2002) wanted to understand the roots of the uniquely American form of capitalism. Compared with European nations, the United States has shown less interest in

providing for the needs of average citizens and has granted greater power to gigantic corporations. Perrow feels the die was pretty much cast by the end of the nineteenth century, resting primarily on Supreme Court decisions in favor of corporations and the experiences of the textile and railroad industries.

- *Diminished Democracy:* Theda Skocpol (2003) turns her attention to something that fascinated Alexis de Tocqueville in his 1840 *Democracy in America:* the grassroots commitment to democracy, which appeared in all aspects of American community life. It almost seemed as though democratic decision making was genetic in the new world, but what happened? Skocpol's analysis of contemporary U.S. culture suggests a "diminished democracy" that cannot be easily explained by the ideologies of either the right or the left.

These examples of comparative and historical research should give you some sense of the potential power of the method. Let's turn now to an examination of the sources and techniques used in this method.

Sources of Comparative and Historical Data

As we saw in the case of existing statistics, there is no end of data available for analysis in historical research. To begin, historians may have already reported on whatever it is you want to examine, and their analyses can give you an initial grounding in the subject, a jumping-off point for more in-depth research.

Most likely you'll ultimately want to go beyond others' conclusions and examine some "raw data" to draw your own conclusions. These data vary, of course, according to the topic under study. When W. I. Thomas and Florian Znaniecki (1918) studied the adjustment process for Polish peasants coming to the United States early in this century, they examined letters written by the immigrants to their families in Poland. (They obtained the letters through newspaper advertisements.) Other researchers have analyzed old diaries. Such personal documents only scratch the surface, however. In discussing procedures for studying the history of family life, Ellen Rothman points to the following sources:

In addition to personal sources, there are public records which are also revealing of family history. Newspapers are especially rich in evidence on the educational, legal, and recreational aspects of family life in the past as seen from a local point of view. Magazines reflect more general patterns of family life; students often find them interesting to explore for data on perceptions and expectations of mainstream family values. Magazines offer several different kinds of sources at once: visual materials (illustrations and advertisements), commentary (editorial and advice columns), and fiction. Popular periodicals are particularly rich in the last two. Advice on many questions of concern to families—from the proper way to discipline children to the economics of wallpaper—fills magazine columns from the early nineteenth century to the present. Stories that suggest common experiences or perceptions of family life appear with the same continuity.

(1981: 53)

Organizations generally document themselves, so if you're studying the development of some organization you should examine its official documents: charters, policy statements, speeches by leaders, and so on. Once, when I was studying the rise of a contemporary Japanese religious group—Sokagakkai—I discovered not only weekly newspapers and magazines published by the group but also a published collection of all the speeches given by the original leaders. With these sources, I could trace changes in recruitment patterns over time. At the outset, followers were enjoined to enroll all the world. Later, the emphasis shifted specifically to Japan. Once a sizable Japanese membership had been established, an emphasis on enrolling all the world returned (Babbie 1966).

Often, official government documents provide the data needed for analysis. To better appreciate the history of race relations in the United States, A. Leon Higginbotham Jr. (1978) examined 200 years of laws and court cases involving race. Himself the first African American appointed a federal judge, Higginbotham found that, rather than protecting African Americans, the law embodied bigotry and oppression. In the earliest court cases, there was considerable ambiguity over whether African Americans

Tips and Tools

Reading and Evaluating Documents

Ron Aminzade and Barbara Laslett

University of Minnesota

The purpose of the following comments is to give you some sense of the kind of interpretive work historians do and the critical approach they take toward their sources. It should help you to appreciate some of the skills historians develop in their efforts to reconstruct the past from residues, to assess the evidentiary status of different types of documents, and to determine the range of permissible inferences and interpretations. Here are some of the questions historians ask about documents:

1. Who composed the documents? Why were they written? Why have they survived all these years? What methods were used to acquire the information contained in the documents?

2. What are some of the biases in the documents and how might you go about checking or correcting them? How inclusive or representative is the sample of individuals, events, and so on, contained in the document? What were the institutional constraints and the general organizational routines under which the document was

prepared? To what extent does the document provide more of an index of institutional activity than of the phenomenon being studied? What is the time lapse between the observation of the events documented and the witnesses' documentation of them? How confidential or public was the document meant to be? What role did etiquette, convention, and custom play in the presentation of the material contained within the document? If you relied solely upon the evidence contained in these documents, how might your vision of the past be distorted? What other kinds of documents might you look at for evidence on the same issues?

3. What are the key categories and concepts used by the writer of the document to organize the information presented? What selectivities or silences result from these categories of thought?

4. What sorts of theoretical issues and debates do these documents cast light on? What kinds of historical and/or sociological questions do they help to answer? What sorts of valid inferences can one make from the information contained in these documents? What sorts of generalizations can one make on the basis of the information contained in these documents?

---

were indentured servants or, in fact, slaves. Later court cases and laws clarified the matter—holding African Americans to be something less than human.

The sources of data for historical analysis are too extensive to cover even in outline here, though the examples we've looked at should suggest some ideas. Whatever resources you use, however, a couple of cautions are in order.

As we saw in the case of existing statistics, you can't trust the accuracy of records—official or unofficial, primary or secondary. Your protection lies in replication: In the case of historical research, that means corroboration. If several sources point to the same set of "facts," your confidence in them might reasonably increase.

At the same time, you need always be wary of bias in your data sources. If all your data on the development of a political movement are taken from the movement itself, you're unlikely to gain a well-rounded view of it. The diaries of well-to-do gentry of the Middle Ages may not give you an accurate view of life in general during those times. Where possible, obtain data from a variety of sources representing different points of view.

As Ron Aminzade and Barbara Laslett indicate in the Tips and Tools box, "Reading and Evaluating Documents," there is an art to knowing how to regard such documents and what to make of them.

Incidentally, the critical review that Aminzade and Laslett urge for the reading of historical documents is useful in many areas of your life besides the pursuit of comparative and historical research. Consider applying some of their questions to presidential press conferences, advertising, or (gasp) college textbooks. None of these offers a direct view of reality; all have human authors and human subjects.

Analytic Techniques

The analysis of comparative and historical data is another large subject that I can't cover exhaustively here. Moreover, because comparative and historical research is usually a qualitative method, there are no easily listed steps to follow in the analysis of historical data. Nevertheless, a few comments are in order.

Max Weber used the German term *verstehen*—"understanding"—in reference to an essential

quality of social research. He meant that the researcher must be able to take on, mentally, the circumstances, views, and feelings of those being studied, so that the researcher can interpret their actions appropriately. Certainly this concept applies to comparative and historical research. The researcher's imaginative understanding is what breathes life and meaning into the evidence being analyzed.

The comparative and historical researcher must find patterns among the voluminous details describing the subject matter of study. Often, this takes the form of what Weber called "ideal types": conceptual models composed of the essential characteristics of social phenomena. For example, Weber himself did considerable

research on bureaucracy. Having observed numerous actual bureaucracies, Weber ([1925] 1946) detailed those qualities essential to bureaucracies in general: jurisdictional areas, hierarchically structured authority, written files, and so on. Weber did not merely list those characteristics common to all the actual bureaucracies he observed. Rather, to create a theoretical model of the "perfect" (ideal type) bureaucracy, he needed to understand fully the essentials of bureaucratic operation. Figure 11-5 offers a more recent illustration of some positive and negative aspects of bureaucracy as a general social phenomenon.

Often, comparative and historical research is informed by a particular theoretical paradigm. Thus, Marxist scholars may undertake historical
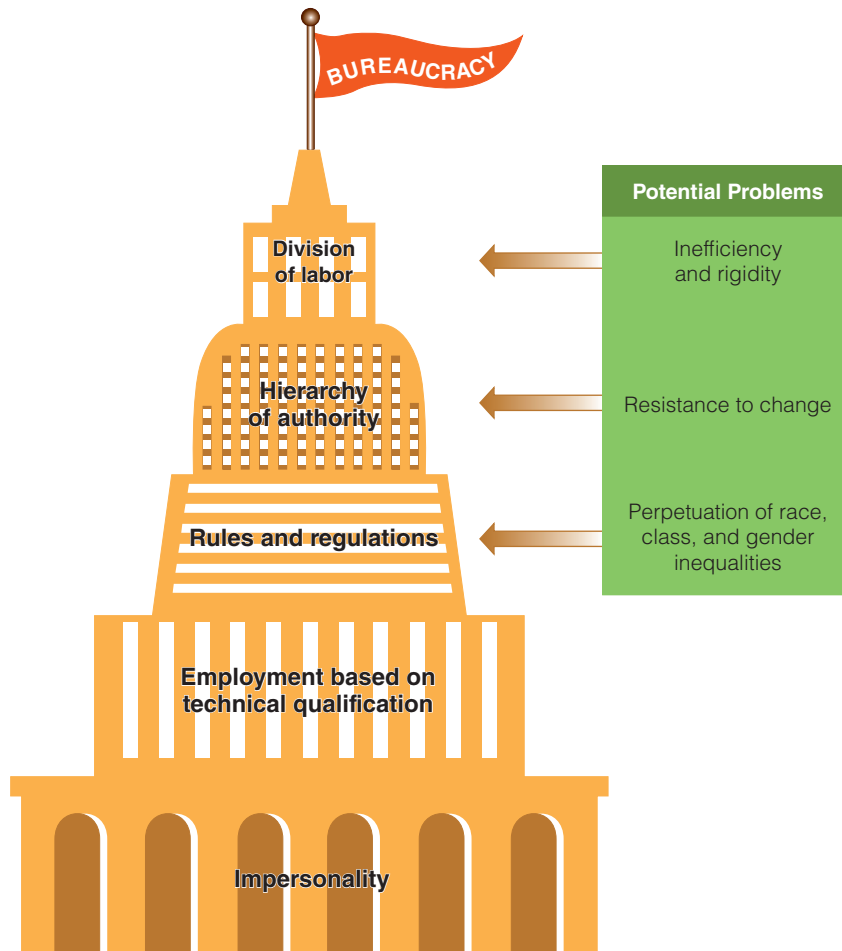

FIGURE 11-5

Some Positive and Negative Aspects of Bureaucracy

Source: Diana Kendall, *Sociology in Our Times*, 5th ed. (Belmont, CA: Wadsworth, © 2005). Used by permission.

analyses of particular situations—such as the history of Latinos and Latinas in the United States—to determine whether they can be understood in terms of the Marxist version of conflict theory. Sometimes, comparative and historical researchers attempt to replicate prior studies in new situations—for example, doing follow-up replications of Weber's studies of religion and economics.

Although comparative and historical research is often regarded as a qualitative rather than quantitative technique, this is by no means necessary. Historical analysts sometimes use time-series data to monitor changing conditions over time, such as data on population, crime rates, unemployment, infant mortality rates, and so forth. The analysis of such data sometimes requires sophistication, however. For example, Larry Isaac and Larry Griffin (1989) discuss the uses of a variation on regression techniques (see Chapter 16) in determining the meaningful breaking points in historical processes, as well as for specifying the periods within which certain relationships occur among variables. Criticizing the tendency to regard history as a steadily unfolding process, the authors focus their attention on the statistical relationship between unionization and the frequency of strikes, demonstrating that the relationship has shifted importantly over time.

Isaac and Griffin raise several important issues regarding the relationship among theory, research methods, and the "historical facts" they address. Their analysis, once again, warns against the naive assumption that history as documented necessarily coincides with what actually happened.

Unobtrusive Online Research

Since this is the final chapter on methods of data collection, it might be useful to review some of the ways in which online data are being used for unobtrusive social research. We've seen the wealth of data sources online, but online processes themselves can be the subject of study.

We defined *big data* in Chapter 2, and it is a term you are likely to hear increasingly in the future. Recall that it refers to the gigantic data sets being automatically compiled from online

activity. The most notorious of these in recent times has been the National Security Agency's (NSA) compilation and analysis of phone, e-mail, and other communications. Prior to the development of massive data-storage capacities, this feat would have been unimaginable.

Big data are also in common use by commercial enterprises such as Google, Amazon, and many retailers. Have you ever noticed that after you've read an article online about cocker spaniels, you may begin receiving e-mail advertising pet products, and the marginal ads on websites you visit in the future may feature similar commodities. Social media such as Facebook and Twitter are another source of big data.

The rise of social media has both impacted society and opened new avenues for unobtrusive social research. For example, Rachel Gong (2011) examined the extent to which the use of social media might affect political success. Noting the general conclusion that President Barack Obama's electoral success rested heavily on his use of electronic media, Gong tested the breadth of this effect—examining the utility of social media in a society such as Malaysia, where the conventional media were state controlled. She found that parliamentary candidates whose campaign maintained a Weblog were far more successful than those who did not have a presence on social media; this was especially true for candidates running in opposition to the status quo.

The importance of social media for politics and other aspects of modern life is not simply a matter of whether it is used but *how.* Noah Smith and colleagues at Carnegie Mellon University are developing several computer programs for analyzing natural language. One of their special interests is the analysis of Twitter communications, and they have used those analyses to explain political positions in Congress and to predict National Football League outcomes, among other things. Katy Steinmetz (2013) has reviewed their work and you can learn more at Noah's Ark (www.ark.cs.cmu.edu).

Increasingly, social research journals are addressing these emerging methodologies. For example, the *International Journal of Social Research Methodology* devotes an entire May 2013 special issue to the topic of digital social research. A review of the contents shows the range and

potential of sociological research in this area, with articles examining new techniques and their implications for social science.

Robert Ackland and Rachel Gibson (2013) examined differences in the way 100 political parties in six countries used their websites for different kinds of communications. Some of the hyperlinks on political websites directed users to candidate or issue websites that reflect the values of the party, while others aimed at discrediting opponents. Still other hyperlinks sought to enhance the party by associating it with larger or more-established groups. Among other things, the researchers found left-wing parties using hyperlinks to demonstrate an international orientation and an affinity for the nonprofit sector. Right-wing parties, by contrast, used hyperlinks to show their affinity for business and the commercial sector and for same-country groups (2013: 241).

Rob Procter, Farida Vis, and Alex Voss (2013) set about to analyze the 2011 London anti-austerity riots by using 54 Twitter hashtags to identify 2.6 million Tweets relating to various aspects of the riots. This enormous mass of data allowed them to monitor the rise and fall of a rumor that the London Eye and Big Ben were on fire, for example. The sheer volume of data required them to invent new techniques for content analysis.

Adam Edwards and colleagues (2013) point to an interesting distinction for social media research in comparison with other social science techniques. Some methods, such as survey research, permit a broad viewing (e.g., society-wide) of social life at a specified time. Participant observation and other methods allow for an immediate scrutiny of social processes but with a much narrower view. Some of the digital research into social media allows a dynamic observation of process—as it is happening—on a scale as broad as a national survey. Overall, the authors do not foresee conventional methods being replaced by digital research, but they do anticipate new potentials for studying human social behavior with these emerging methodologies.

The variety of online sources and techniques opens new possibilities for approaching research questions from several directions. If each independent approach produces the same conclusion, we can be more confident of our findings.

Seth Stephens-Davidowitz (2013) offered an intriguing example when he sought to estimate the percentage of American men who are gay. Noting surveys may ask for this information, he worried that many gay men would not report their sexuality, especially in states and communities that were intolerant to gays.

Some of the data sources he examined online were Facebook profiles indicating preference for same-sex partners. He found fewer such profiles for Facebook members in states intolerant to gays. To accommodate for the possibility that gay men might move to more-tolerant states, Stephens-Davidowitz focused his attention on high school students, reasoning they would be less able to pack up and move.

Stephens-Davidowitz also obtained data from match.com, an online dating site, and Craigslist. Using Google, he was able to examine searches for gay porn sites and married women searching for answers to the question: "Is my husband gay?" Each of these approaches offered estimates of the percentage of openly gay men and the percentage still in the closet.

Ethics and Unobtrusive Measures

The use of unobtrusive measures avoids many of the ethical issues we've discussed in connection with other data-collection techniques, but if you reflect on the general principles we've covered, I think you'll see that there are potential risks to guard against.

For example, the general principle of confidentiality may be relevant in some projects. Let's suppose you want to examine an immigrant subculture through a content analysis of letters written back to the old country, as was the case in the Thomas and Znaniecki (1918) study of Polish peasants, mentioned earlier in the chapter. To begin, you should obtain those letters legally and ethically (no getting a government agency to intercept the letters for you), and you need to protect the privacy of the letter writers and recipients.

As with all other research techniques, you're obliged to collect data, analyze them, and report your findings honestly, with the purpose of discovering what is so, rather than attempting to support a favored hypothesis or personal agenda.

While it may be easy to agree with such a principle, you're likely to find it somewhat more difficult to apply when you actually conduct research. Your ethical sensibilities will be more challenged by the vast gray areas than by those of black and white.

MAIN POINTS

Introduction

- Unobtrusive measures are ways of studying social behavior without affecting it in the process.

Content Analysis

- Content analysis is a social research method appropriate for studying human communications through social artifacts. Researchers can use it to study not only communication processes but other aspects of social behavior as well.

- Common units of analysis in content analysis include elements of communications—words, paragraphs, books, and so forth. Standard probability-sampling techniques are sometimes appropriate in content analysis.

- Content analysis involves coding—transforming raw data into categories based on some conceptual scheme. Coding may attend to both manifest and latent content. The determination of latent content requires judgments by the researcher.

- Both quantitative and qualitative techniques are appropriate for interpreting content analysis data.

- The advantages of content analysis include economy, safety, and the ability to study processes occurring over a long time. Also, it is unobtrusive. Its disadvantages are that it is limited to recorded communications and can raise issues of reliability and validity.

Analyzing Existing Statistics

- A variety of government and nongovernment agencies provide aggregate statistical data for studying aspects of social life.

- Problems of validity in the analysis of existing statistics can often be handled through logical reasoning and replication.

- Existing statistics often have problems of reliability, so they must be used with caution.

Comparative and Historical Research

- Social scientists use comparative and historical methods to discover patterns in the histories of different cultures.

- Although often regarded as a qualitative method, comparative and historical research can make use of quantitative techniques, such as the analysis of time-series data.

Unobtrusive Online Research

- Social media and other online activities have opened up a new realm of possibilities for unobtrusive research.

- Researchers are developing new techniques for analyzing online content.

Ethics and Unobtrusive Measures

- Sometimes even unobtrusive measures can raise the possibility of violating subjects' privacy.

- The general principles of honest observation, analysis, and reporting apply to all research techniques.

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.

coding	latent content
comparative and historical research	manifest content
content analysis	unobtrusive research

PROPOSING SOCIAL RESEARCH: UNOBTRUSIVE RESEARCH

This chapter has provided an overview of three major types of unobtrusive research: content analysis, analyzing existing statistics, and comparative and historical research. While existing statistics represent, by their nature, a quantitative method, the other two can be done with a qualitative and/or quantitative approach. In this exercise, you need to identify which method and orientation you'll use. If you're doing these exercises in order to understand the topics of the book better, you could try your hand at each of these methods.

You need to describe the data you'll use and detail anything special about your access to those data. Whether you're studying newspaper editorials, infant mortality rates, or accounts of political revolutions, you'll likely face potential problems of validity and reliability. Unobtrusive methods involve the use of available data, which often offer approximations of the observations you might ideally like to make. For example, you may need to use drug-arrest rates as an approximation of drug-use rates. You should discuss how you'll deal with any such approximations.

REVIEW QUESTIONS AND EXERCISES

1. Outline a content analysis design to determine whether the Republican or the Democratic party is the more supportive of a basic constitutional right such as free speech, freedom of religion, or protection against self-incrimination. Be sure to specify units of analysis and sampling methods. Describe a coding scheme that you could use for the content analysis.

2. Identify an international news story involving a conflict between two nations or cultural groups, such as clashes between Israelis and Palestinians. On the Internet, locate a newspaper report of the event from within each of the countries or cultures involved. Note differences in the way the event is reported. Now, find a report of the event in a newspaper in a third, distant country. (For example, compare reports from the *Jerusalem Post,* the *Palestine Chronicle,* and the *New York Times.*) Does the third report seem to favor one of the two original reports? If so, would you conclude that the third report is biased toward one side or that one of the original reports was simply inaccurate? Explain how and why you reached that conclusion.

3. Using the web, find out how many countries have a higher "expected life expectancy" than the United States does.

4. Max Weber undertook extensive studies of some of the world's major religions. Create an annotated bibliography of his works in this area.

5. On the web, locate the American Sociological Association's section called "Comparative and Historical Sociology." Summarize an article in the section's newsletter.

Evaluation Research

Now you're going to see one of the most rapidly growing uses of social research: the evaluation of social interventions. You'll come away from this chapter able to judge whether social programs have succeeded or failed.

Introduction

You may not be familiar with *Twende na Wakati* (translated: *Let's Go with the Times*), but it was the most popular radio show in Tanzania a few years back. It was a soap opera. The main character, Mkwaju, was a truck driver with some pretty traditional ideas about gender roles and sexuality. By contrast, Fundi Mitindo, a tailor, and his wife, Mama Waridi, had more-modern ideas regarding the roles of men and women, particularly in relation to the issues of overpopulation and family planning.

Twende na Wakati was the creation of Population Communications International (PCI) and other organizations working in conjunction with the Tanzanian government in response to two problems facing that country: (1) a population growth rate over twice that of the rest of the world and (2) an AIDS epidemic particularly heavy along the international truck route, where more than a fourth of the truck drivers and over half the commercial sex workers were found to be HIV positive in 1991. The prevalence of contraceptive use was 11 percent (Rogers et al. 1996: 5–6).

The purpose of the soap opera was to bring about a change in knowledge, attitudes, and practices (KAP) relating to contraception and family planning. Rather than instituting a conventional educational campaign, PCI felt it would be more effective to illustrate the message through entertainment.

Between 1993 and 1995, there were 208 episodes of *Twende na Wakati* aired, aiming at the 67 percent of Tanzanians who listen to the radio. Eighty-four percent of the radio listeners reported listening to the PCI soap opera, making it the most popular show in the country. Ninety percent of the show's listeners recognized Mkwaju, the sexist truck driver, and only 3 percent regarded him as a positive role model. Over two-thirds identified Mama Waridi, a businesswoman, and her tailor husband as positive role models.

Surveys conducted to measure the impact of the show indicated it had affected knowledge, attitudes, and behavior. For example, 49 percent of the married women who listened to the show said they now practiced family planning, compared with only 19 percent of the nonlisteners. There were other impacts:

> Some 72 percent of the listeners in 1994 said that they adopted an HIV/AIDS prevention behavior because of listening to "Twende na Wakati," and this percentage increased to 82 percent in our 1995 survey. Seventy-seven percent of these individuals adopted monogamy, 16 percent began using condoms, and 6 percent stopped sharing razors and/or needles.
>
> *(Rogers et al. 1996: 21)*

We can judge the effectiveness of the soap opera because of a particular form of social science. *Evaluation research* refers to a research purpose rather than a specific method. This purpose is to evaluate the impact of social interventions such as new teaching methods or innovations in parole. Many methods—surveys, experiments, and so on—can be used in evaluation research.

Evaluation research is appropriate whenever some social intervention occurs or is planned. A *social intervention* is an action taken within a social context for the purpose of producing some intended result. In its simplest sense, **evaluation research** is the process of determining whether a social intervention has produced the intended result. Peter Rossi, Mark Lipsey, and Howard Freeman (2002: 4) define it as follows:

> Program evaluation is the use of social research procedures to systematically investigate the effectiveness of social intervention programs. More specifically, evaluation researchers [evaluators] use social research methods to study, appraise, and help improve social programs in all their important aspects, including the diagnosis of the social problems they address, their conceptualization and design, their implementation and administration, their outcomes, and their efficiency.

evaluation research Research undertaken for the purpose of determining the impact of some social intervention, such as a program aimed at solving a social problem.

Evaluation research is probably as old as social research itself. Whenever people have instituted a social reform for a specific purpose, they have paid attention to its actual consequences, even if they have not always done so in a conscious, deliberate, or sophisticated fashion. In recent years, however, the field of evaluation research has become an increasingly popular and active research specialty, as reflected in textbooks, courses, and projects. Moreover, the growth of evaluation research points to a more general trend in the social sciences. As a researcher, you'll likely be asked to conduct evaluations of your own.

In part, the growth of evaluation research reflects social researchers' increasing desire to make a difference in the world. At the same time, we can't discount the influence of (1) an increase in federal requirements that program evaluations must accompany the implementation of new programs and (2) the availability of research funds to fulfill those requirements. In any case, it seems clear that social researchers will be bringing their skills into the real world more than ever before.

This chapter looks at some of the key elements in this form of social research. After considering the kinds of topics commonly subjected to evaluation, we'll move through some of its main operational aspects: measurement, study design, and execution. As you'll see, formulating questions is as important as answering them. Because it occurs within real life, evaluation research has its own problems, some of which we'll examine. Logistical problems arise from evaluation research generally and from its specific, technical procedures. The use of research results also presents certain concerns. As you review reports of program evaluations, you should be especially sensitive to these issues.

Evaluation is a form of applied research—that is, it's intended to have some real-world effect. It will be useful, therefore, to consider whether and how it's actually applied. As you'll see, the obvious implications of an evaluation research project do not necessarily affect real life. They may become the focus of ideological, rather than scientific, debates. They may simply be denied out of hand, for political or other reasons. Perhaps most typically, they may simply be ignored and forgotten, left to collect dust in bookcases across the land.

The chapter continues with a look at a particular resource for large-scale evaluation—social indicators research. This type of research is also a rapidly growing specialty. Essentially it involves the creation of aggregated indicators of the "health" of society, similar to the economic indicators that give diagnoses and prognoses of economies. The chapter then concludes with a look at the special ethical concerns that arise in evaluation research.

Topics Appropriate for Evaluation Research

The topics appropriate for evaluation research are limitless. When the federal government abolished the selective service system (military draft), military researchers began paying special attention to the impact on enlistment. As individual states have liberalized their marijuana laws, researchers have sought to learn the consequences, both for marijuana use and for other forms of social behavior. Do no-fault divorce reforms increase the number of divorces, and do related social problems decrease or increase? Has no-fault automobile insurance really brought down insurance policy premiums? Agencies providing foreign aid also conduct evaluations to determine whether the desired effects were produced. Has the "No Child Left Behind" program improved the quality of education in America? Have "Just Say No" abstinence programs reduced rates of sexual activity and pregnancies among young people? These are the kinds of questions that evaluation research can address.

There are many variations in the intent of evaluation research. **Needs assessment studies** aim to determine the existence and extent of problems, typically among a segment of the population, such as the elderly. **Cost-benefit studies** determine whether the results of a program can be

needs assessment studies Studies that aim to determine the existence and extent of problems, typically among a segment of the population, such as the elderly.

cost-benefit studies Studies that determine whether the results of a program can be justified by its expense (both financial and other).

justified by its expense (both financial and other). **Monitoring studies** provide a steady flow of information about something of interest, such as crime rates or the outbreak of an epidemic. Sometimes the monitoring involves incremental interventions. Read this description of "adaptive management" by the Nature Conservancy, a public-interest group seeking to protect natural areas:

> First, partners assess assumptions and set management goals for the conservation area. Based on this assessment, the team takes action, then monitors the environment to see how it responds. After measuring results, partners refine their assumptions, goals and monitoring regimen to reflect what they've learned from past experiences. With refinements in place, the entire process begins again.
>
> *(2005: 3)*

Much of evaluation research is referred to as **program evaluation** or **outcome assessment**: the determination of whether a social intervention is producing the intended result. Here's an example.

Some years ago, a project evaluating the nation's drivers' education programs, conducted by the National Highway and Transportation Safety Administration (NHTSA), stirred up a controversy. Philip Hilts (1981: 4) reported on the study's findings:

> For years the auto insurance industry has given large insurance discounts for children who take drivers' education courses, because statistics show that they have fewer accidents.
>
> The preliminary results of a new major study, however, indicate that drivers' education does not prevent or reduce the incidence of traffic accidents at all.

Based on an analysis of 17,500 young people in DeKalb County, Georgia (including Atlanta),

monitoring studies Studies that provide a steady flow of information about something of interest, such as crime rates or the outbreak of an epidemic.

program evaluation/outcome assessment The determination of whether a social intervention is producing the intended result.

the preliminary findings indicated that students who took drivers' education had just as many accidents and traffic violations as those who didn't take it. The study also seemed to reveal some subtle aspects of driver training.

First, it suggested that the apparent impact of drivers' education was largely a matter of self-selection. The kind of students who took drivers' education were less likely to have accidents and traffic violations—with or without driver training. Students with high grades, for example, were more likely to sign up for driver training, and they were also less likely to have accidents.

More startling, however, was the suggestion that driver-training courses may have actually increased traffic accidents! The existence of drivers' education may have encouraged some students to get their licenses earlier than if there were no such courses. In a study of ten Connecticut towns that discontinued driver training, about three-fourths of those who probably would have been licensed through their classes delayed getting licenses until they were 18 or older (Hilts 1981: 4).

As you might imagine, these results were not well received by those most closely associated with driver training. This matter was complicated, moreover, by the fact that the NHTSA study was also evaluating a new, more intensive training program—and the preliminary results showed that the new program was effective.

Here's a very different example of evaluation research. Rudolf Andorka, a Hungarian sociologist, had been particularly interested in his country's shift to a market economy. Even before the dramatic events in Eastern Europe in 1989, Andorka and his colleagues had been monitoring the nation's "second economy"—jobs pursued outside the socialist economy. Their surveys followed the rise and fall of such jobs and examined their impact within Hungarian society. One conclusion was that "the second economy, which earlier probably tended to diminish income inequalities or at least improved the standard of living of the poorest part of the population, in the 1980s increasingly contributed to the growth of inequalities" (Andorka 1990: 111).

Whereas evaluation research is basically a matter of discovering whether social interventions make a difference, it is not surprising that it is sometimes coupled with the intentions of

participatory action research (PAR), discussed in Chapter 10. Since PAR has been particularly strong among Australian researchers, it's not surprising to find Wayne Miller and June Lennie (2005) speaking of "empowerment evaluation" to characterize their assessment of a national school-breakfast program. They say that this approach aims to include all types of stakeholders—staff, funders, members of the community—in the design and execution of the evaluation. And in the process, they intend that evaluation and improvement will "become a normal part of planning and managing programs." (2005: 18)

As you can see, the questions appropriate to evaluation research are of great practical significance: Jobs, programs, and investments as well as beliefs and values are at stake. Let's now examine how these questions are answered—how evaluations are conducted.

Formulating the Problem: Issues of Measurement

Several years ago, I headed an institutional research office that conducted research directly relevant to the operation of the university. Often, we were asked to evaluate new programs in the curriculum. The following description is fairly typical of the problem that arose in that context, and it points to one of the key barriers to good evaluation research.

Faculty members would appear at my office to say they'd been told by the university administration to arrange for an evaluation of the new program they had permission to try. This points to a common problem: Often the people whose programs are being evaluated aren't thrilled at the prospect. For them, an independent evaluation threatens the survival of the program and perhaps even their jobs.

The main problem I want to introduce, however, has to do with the purpose of the intervention to be evaluated. The question "What is the intended result of the new program?" often produced a vague response such as "Students will get an in-depth and genuine understanding of mathematics, instead of simply memorizing methods of calculations." Fabulous! And how could we measure that "in-depth and genuine understanding"? Often, I was told that the program aimed at

producing something that could not be measured by conventional aptitude and achievement tests. No problem there; that's to be expected when we're innovating and being unconventional. What would be an unconventional measure of the intended result? Sometimes this discussion came down to an assertion that the effects of the program would be "unmeasurable."

There's the common rub in evaluation research: measuring the "unmeasurable." Evaluation research is a matter of finding out whether something is there or not there, whether something happened or didn't happen. To conduct evaluation research, we must be able to operationalize, observe, and recognize the presence or absence of what is under study.

Often, outcomes can be derived from published program documents. Thus, when Edward Howard and Darlene Norman (1981) evaluated the performance of the Vigo County Public Library (VCPL) in Indiana, they began with the statement of purpose previously adopted by the library's Board of Trustees.

> To acquire by purchase or gift, and by recording and production, relevant and potentially useful information that is produced by, about, or for the citizens of the community;
>
> To organize this information for efficient delivery and convenient access, furnish the equipment necessary for its use, and provide assistance in its utilization; and
>
> To effect maximum use of this information toward making the community a better place in which to live through aiding the search for understanding by its citizens.
>
> *(1981: 306)*

As the researchers said, "Everything that VCPL does can be tested against the Statement of Purpose." They then set about creating operational measures for each of the purposes.

Although "official" purposes of interventions are often the key to designing an evaluation, they may not always be sufficient. Anna-Marie Madison (1992: 38), for example, warns that programs designed to help disadvantaged minorities do not always reflect what the proposed recipients of the aid may need and desire:

> The cultural biases inherent in how middle-class white researchers interpret the experiences of low-income minorities may lead to

erroneous assumptions and faulty proposi-
tions concerning causal relationships, to
invalid social theory, and consequently to
invalid program theory. Descriptive theories
derived from faulty premises, which have
been legitimized in the literature as existing
knowledge, may have negative consequences
for program participants.

In setting up an evaluation, then, research-
ers must pay careful attention to issues of mea-
surement. Let's take a closer look at the types of
measurements that evaluation researchers must
deal with.

Specifying Outcomes

As I've already suggested, a key variable for eval-
uation researchers to measure is the outcome, or
what is called the *response variable.* If a social pro-
gram is intended to accomplish something, we
must be able to measure that something. If we
want to reduce prejudice, we need to be able to
measure prejudice. If we want to increase marital
harmony, we need to be able to measure that.

It's essential to achieve agreements on defi-
nitions in advance:

> The most difficult situation arises when there
> is disagreement as to standards. For example,
> many parties may disagree as to what defines
> serious drug abuse—is it defined best as 15%
> or more of students using drugs weekly,
> 5% or more using hard drugs such as cocaine
> or PCP monthly, students beginning to use
> drugs as young as seventh grade, or some
> combination of the dimensions of rate of use,
> nature of use, and age of user? . . . Applied
> researchers should, to the degree possible,
> attempt to achieve consensus from research
> consumers in advance of the study (e.g.,
> through advisory groups) or at least ensure
> that their studies are able to produce data
> relevant to the standards posited by all po-
> tentially interested parties.
>
> *(Hedrick, Bickman, and Rog 1993: 27)*

In some cases you may find that the definitions
of a problem and a sufficient solution are defined
by law or by agency regulations; if so, you must
be aware of such specifications and accommo-
date them. Moreover, whatever the agreed-on
definitions, you must also achieve agreement on

how the measurements will be made. Because
there are different possible methods for estimating
the percentage of students "using drugs weekly,"
for example, you'd have to be sure that all the
parties involved understood and accepted the
method(s) you've chosen.

Or on the other side of the coin, Yuet Wah
Cheung (2009) used "drug-free weeks" as the
dependent variable in his evaluation of drug-
treatment programs in Hong Kong. This lon-
gitudinal study examined the role of positive
and negative "social capital" in determining
success or failure. *Positive social capital* included
degree of family support and support from non-
drug-using friends, while *negative social capital*
included stressful events and association with
drug-using friends. Cheung found, for example,
that if recovering drug users were able to es-
tablish networks of supportive, non-drug-using
friends, this made it less likely that they would
revert to associating with their old network of
drug users.

In the case of the Tanzanian soap opera,
there were several outcome measures. In part,
the purpose of the program was to improve
knowledge about both family planning and
AIDS. Thus, for example, one show debunked
the belief that the AIDS virus was spread by mos-
quitoes and could be avoided by the use of insect
repellant. Studies of listeners showed a reduction
in that belief (Rogers et al. 1996: 21).

PCI also wanted to change Tanzanian at-
titudes toward family size, gender roles, HIV/
AIDS, and other related topics; the research in-
dicated that the show had affected these as well.
Finally, the program aimed at affecting behavior.
We've already seen that radio listeners reported
changing their behavior with regard to AIDS
prevention. They reported a greater use of family
planning as well. However, because there's
always the possibility of a gap between what
people say they do and what they actually do,
the researchers sought independent data to
confirm their conclusions.

Tanzania's national AIDS-control program
had been offering condoms free of charge to citi-
zens. In the areas covered by the soap opera, the
number of condoms given out increased sixfold
between 1992 and 1994. This far exceeded the
increase of 1.4 times in the control area, where
broadcasters did not carry the soap opera.

Measuring Experimental Contexts

Measuring the dependent variables that are directly involved in the experimental program is only a beginning. As Henry Riecken and Robert Boruch (1974: 120–21) point out, it's often appropriate and important to measure those aspects of the context of an experiment researchers think might affect the experiment. Though external to the experiment itself, some variables may affect it.

Suppose, for example, that you were conducting an evaluation of a program aimed at training unskilled people for employment. The primary outcome measure would be their success at gaining employment after completing the program. You would, of course, observe and calculate the subjects' employment rate, but you should also determine what has happened to the employment/unemployment rates of society at large during the evaluation. A general slump in the job market should be taken into account in assessing what might otherwise seem a relatively low employment rate for subjects. Or, if all the experimental subjects get jobs following the program, you should consider any general increase in available jobs. Combining complementary measures with proper control-group designs should allow you to pinpoint the effects of the program you're evaluating.

Specifying Interventions

Besides making measurements relevant to the outcomes of a program, researchers must measure the program intervention—the experimental stimulus. In part, this measurement will be handled by the assignment of subjects to experimental and control groups, if that's the research design. Assigning a person to the experimental group is the same as scoring that person "yes" on the stimulus, and assignment to the control group represents a score of "no." In practice, however, it's seldom that simple.

Let's stick with the job-training example. Some people will participate in the program; others will not. But imagine for a moment what job-training programs are probably like. Some subjects will participate fully; others will miss a lot of sessions or fool around when they are present. So you may need measures of the extent or quality of participation in the program. If the program is effective, you should find that those who participated fully have higher employment rates than do those who participated less.

Other factors may further confound the administration of the experimental stimulus. Suppose we're evaluating a new form of psychotherapy designed to cure sexual impotence. Several therapists administer it to subjects composing an experimental group. We plan to compare the recovery rate of the experimental group with that of a control group, which receives some other therapy or none at all. It may be useful to include the names of the therapists treating specific subjects in the experimental group, because some may be more effective than others. If this turns out to be the case, we must find out why the treatment worked better for some therapists than for others. What we learn will further develop our understanding of the therapy itself.

The Tips and Tools box, "Positive Deviance," offers an alternative view for designing an intervention.

Specifying the Population

In evaluating an intervention, it's important to define the population of possible subjects for whom the program is appropriate. Ideally, all or a sample of appropriate subjects will then be assigned to experimental and control groups as warranted by the study design. Defining the population, however, can itself involve specifying measurements. If we're evaluating a new form of psychotherapy, for example, it's probably appropriate for people with mental problems. But how will "mental problems" be defined and measured? The job-training program mentioned previously is probably intended for people who are having trouble finding work, but what counts as "having trouble"?

Beyond defining the relevant population, then, the researcher should make fairly precise measurements of the variables considered in the definition. For example, even though the randomization of subjects in the psychotherapy study would ensure an equal distribution of those with mild and those with severe mental problems into the experimental and control groups, we'd need to keep track of the relative severity of different subjects' problems in case the therapy turns out to be effective only for

Tips and Tools

Positive Deviance

In his examination of "positive deviance," Arvind Singhal (2011, 2013) points to an implicit idea about how we design social interventions. He uses an example of a rural village in Vietnam, where juvenile malnutrition is a chronic problem. Suppose we want to design a program to solve the problem. The usual approach, Singhal suggests, is to identify examples of the problem—malnourished kids—and use what we learn about their plight to design a solution.

As an alternative, Singhal suggests that we might also look for cases that deviate from the norm positively—children who are not malnourished—and ask why not. Some cases will have obvious explanations. For example, children in the wealthiest family in the village will probably not suffer malnutrition. There may be other cases with obvious explanations, but some will represent more of a puzzle. He reports one poor family in which the children were quite healthy.

When researchers, Jerry and Monica Sternin, studied the family, they learned several ways in which the positive deviance [PD] family differed from others in the village. Singhal (2011: 198—99) summarizes:

- Family members collected tiny shrimps and crabs from paddy fields, adding them to their children's meals. These foods are rich in protein and minerals.
- Family members added greens of sweet potato plants to their children's meals. These greens are rich in beta carotene, and other essential micronutrients, e.g., iron and calcium.

- Interestingly, these foods were accessible to everyone, but most community members believed the foods were inappropriate for young children.

Further,

- PD mothers were feeding their children three to four times a day, rather than the customary twice a day.
- PD mothers were actively feeding their children, making sure there was no food wasted.
- PD mothers washed the hands of the children before and after they ate.

This approach to social change fits well within the context of our discussion of participant action research. Just as physicians feel they know more about their patients' bodies than the patients themselves, social researchers can fall into the trap of discounting what the subjects of their study know about their own situations. When we do that, we may miss a powerful resource for understanding and improving social life.

Sources: Arvind Singhal. 2011. "Turning Diffusion of Innovations Paradigm on Its Head." Pp. 193—205 in *The Diffusion of Innovations: A Communication Science Perspective*, edited by Arun Vishwanath and George A. Barnett. (New York: Peter Lang); Arvind Singhal. 2013. "The Value of Positive Deviations." *Monthly Developments Magazine*, June, pp. 17—20.

those with mild disorders. Similarly, we should measure such demographic variables as sex, age, race, and so forth in case the therapy works only for women, the elderly, or some other group.

New versus Existing Measures

In providing for the measurement of these different kinds of variables, the researcher must continually choose whether to create new measures or use ones already devised by others. If a study addresses something that's never been measured before, the choice is easy. If it addresses something that others have tried to measure, the researcher will need to evaluate the relative worth of various existing measurement devices in terms of her or his specific research situations and purpose. Recall that this is a general issue in social research that applies well beyond evaluation

research. Let's briefly compare creating new measures and using existing ones.

Creating measurements specifically for a study can offer greater relevance and validity than using existing measures would. If the psychotherapy we're evaluating aims at a specific aspect of recovery, we can create measures that pinpoint that aspect. We might not be able to find any standardized psychological measures that hit that aspect right on the head. However, creating our own measure will cost us the advantages to be gained from using preexisting measures. Creating good measures takes time and energy, both of which could be saved by adopting an existing technique. Of greater scientific significance, measures that have been used frequently by other researchers carry a body of possible comparisons that might be important to our evaluation. If the experimental therapy raises scores by an average of ten points

on a standardized test, we'll be in a position to compare that therapy with others that had been evaluated using the same measure. Finally, measures with a long history of use usually have known degrees of validity and reliability, but newly created measures will require pretesting or will be used with considerable uncertainty.

Operationalizing Success/Failure

Potentially one of the most taxing aspects of evaluation research is determining whether the program under review succeeded or failed. The purpose of a foreign language program may be to help students better learn the language, but how much better is enough? The purpose of a conjugal visit program at a prison may be to raise morale, but how high does morale need to be raised to justify the program?

As you may anticipate, clear-cut answers to questions like these almost never arrive. This dilemma has surely been the source of what is generally called *cost-benefit analysis*. How much does the program cost in relation to what it returns in benefits? If the benefits outweigh the cost, keep the program going. If the reverse, junk it. That's simple enough, and it seems to apply in straightforward economic situations: If it cost you $20 to produce something and you can sell it for only $18, there's no way you can make up the difference in volume.

Unfortunately, the situations faced by evaluation researchers are seldom amenable to straightforward economic accounting. The foreign language program may cost the school district $100 per student, and it may raise students' performances on tests by an average of 15 points. Because the test scores can't be converted into dollars, there's no obvious ground for weighing the costs and benefits.

Sometimes, as a practical matter, the criteria of success and failure can be handled through competition among programs. If a different foreign language program costs only $50 per student and produces an increase of 20 points in test scores, it will undoubtedly be considered more successful than the first program—assuming that test scores are seen as an appropriate measure of the purpose of both programs and the less expensive program has no unintended negative consequences.

When Scott Connolly, Katie Elmore, and Wendi Stein (2008) undertook a qualitative evaluation of a Jamaican radio drama, *Outta Road*, designed for youth, they utilized focus groups, in-depth interviews, and exercises in which respondents drew sketches to illustrate their answers. The researchers described their aims thusly:

> The purpose of the study was to assess how listeners to the program engaged with the program and to what extent they found personal meaning and were influenced by the educational messages and themes in the drama.
>
> Unlike a quantitative evaluation, this report does not attempt to generalize the findings to all *Outta Road* youth listeners in Jamaica. The findings do, however, provide rich verbal and visual insights into how the program was incorporated into the lives of participants, what personal meaning they derived from the content, and through reflection how youth listeners internalized the key messages from the drama.
>
> *(2008: 2)*

Ultimately, the criteria of success and failure are often a matter of agreement. The people responsible for the program may commit themselves in advance to a particular outcome that will be regarded as an indication of success. If that's the case, all you need to do is make absolutely certain that the research design will measure the specified outcome. I mention this obvious requirement simply because researchers sometimes fail to meet it, and there's little or nothing more embarrassing than that. So, for example, it is agreed that higher scores on the SAT is the desired result of an educational capstone program, you should ask "how high" and make certain your research design includes SAT scores.

In summary, researchers must take measurement quite seriously in evaluation research, carefully determining all the variables to be measured and getting appropriate measures for each. However, such decisions are typically not purely scientific ones, as we've seen. Evaluation researchers often must work out their measurement strategy with the people responsible for the program being evaluated. It usually doesn't make sense to determine whether a program achieves

Outcome X when its purpose is to achieve Outcome Y. (Realize, however, that evaluation designs sometimes have the purpose of testing for unintended consequences.)

There is a political aspect to these choices, also. Because evaluation research often affects other people's professional interests—their pet program may be halted, or they may be fired or lose professional standing—the results of evaluation research are often argued about.

Let's turn now to some of the research designs commonly employed by evaluators.

Types of Evaluation Research Designs

As I noted at the start of this chapter, evaluation research is not itself a method, but rather one application of social research methods. As such, it can involve any of several research designs. Here we'll consider three main types of research design that are appropriate for evaluations: experimental designs, quasi-experimental designs, and qualitative evaluations.

Experimental Designs

Many of the experimental designs introduced in Chapter 8 can be used in evaluation research. By way of illustration, let's see how the classical experimental model might be applied to our evaluation of a new psychotherapy treatment for sexual impotence.

In designing our evaluation, we should begin by identifying a population of patients appropriate for the therapy. This identification might be made by researchers experimenting with the new therapy. Let's say we're dealing with a clinic that already has 100 patients being treated for sexual impotence. We might take that group and the clinic's definition of sexual impotence as a starting point, and we should maintain any existing assessments of the severity of the problem for each specific patient.

For purposes of evaluation research, however, we would need to develop a more specific measure of impotence. Maybe it would involve whether patients have sexual intercourse at all within a specified time, how often they have intercourse, or whether and how often they reach orgasm. Alternatively, the outcome measure might be based on the assessments of independent therapists not involved in the therapy who interview the patients later. In any event, we would need to agree on the measures to be used.

In the simplest design, we would assign the 100 patients randomly to experimental and control groups; the former would receive the new therapy, and the latter would be taken out of therapy altogether during the experiment. Because ethical practice would probably prevent withdrawing therapy altogether from the control group, however, it's more likely that the control group would continue to receive their conventional therapy.

Having assigned subjects to the experimental and control groups, we would need to agree on the length of the experiment. Perhaps the designers of the new therapy feel it ought to be effective within two months, and an agreement could be reached. The duration of the study doesn't need to be rigid, however. One purpose of the experiment and evaluation might be to determine how long it actually takes for the new therapy to be effective. Conceivably, then, an agreement could be struck to measure recovery rates weekly, say, and let the ultimate length of the experiment rest on a continual review of the results.

Let's suppose the new therapy involves showing pornographic movies to patients. We'd need to specify that stimulus. How often would patients see the movies, and how long would each session be? Would they see the movies in private or in groups? Should therapists be present? Perhaps we should observe the patients while the movies are being shown and include our observations as well as the measurements of the experimental stimulus. Do some patients watch the movies eagerly but others look away from the screen? We'd have to ask these kinds of questions and create specific measurements to address them.

Having thus designed the study, all we have to do is "roll 'em." The study is set in motion, the observations are made and recorded, and the mass of data is accumulated for analysis. Once the study has run its course, we can determine whether the new therapy had its intended—or perhaps some unintended—consequences. We can tell whether the movies were most effective

for mild problems or severe ones, whether they worked for young subjects but not older ones, and so forth.

This simple illustration shows how the standard experimental designs presented in Chapter 8 can be used in evaluation research. Many, perhaps most, of the evaluations reported in the research literature don't look exactly like this illustration, however. Because it's nested in real life, evaluation research often calls for quasi-experimental designs. Let's see what this means.

Quasi-Experimental Designs

Quasi experiments are distinguished from "true" experiments primarily by the lack of random assignment of subjects to an experimental and a control group. In evaluation research, it's often impossible to achieve such an assignment of subjects. Rather than forgo evaluation altogether, researchers sometimes create designs that give some evaluation of the program in question. This section describes some of these designs.

Time-Series Designs

To illustrate the **time-series design**—which involves measurements taken over time—I'll begin by asking you to assess the meaning of some hypothetical data. Suppose I come to you with what I say is an effective technique for getting students to participate in classroom sessions of a course I'm teaching. To prove my assertion, I tell you that on Monday only four students asked questions or made a comment in class; on Wednesday I devoted the class time to an open discussion of a controversial issue raging on campus; and on Friday, when we returned to the subject matter of the course, eight students asked questions or made comments. In other words, I contend, the discussion of a controversial issue on Wednesday has doubled classroom participation. This simple set of data is presented graphically in Figure 12-1.

Have I persuaded you that the open discussion on Wednesday has had the consequence I claim for it? Probably you'd object that my data don't prove the case. Two observations (Monday and Friday) aren't really enough to prove anything. Ideally I should have had two classes, with students assigned randomly to each, held an open discussion in only one, and then compared the two on Friday. But I don't have two classes with
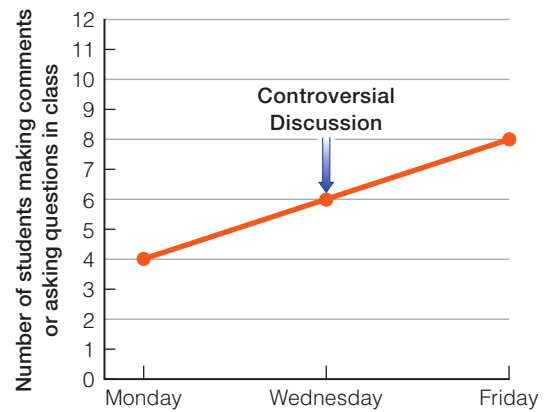
FIGURE 12-1

Two Observations of Class Participation: Before and After an Open Discussion
© Cengage Learning®

random assignment of students. Instead, I've been keeping a record of class participation throughout the semester for the one class. This record allows you to conduct a time-series evaluation.

Figure 12-2 presents three possible patterns of class participation over time, both before and after the open discussion on Wednesday. Which of these patterns would give you some confidence that the discussion had the impact I contend it had?

If the time-series results looked like the first pattern in Figure 12-2, you'd probably conclude that the process of greater class participation had begun on the Wednesday before the discussion and had continued, unaffected, after the day devoted to the discussion. The long-term data suggest that the trend would have occurred even without the discussion on Wednesday. The first pattern, then, contradicts my assertion that the special discussion increased class participation.

The second pattern contradicts my assertion by indicating that class participation has been bouncing up and down in a regular pattern throughout the semester. Sometimes it increases from one class to the next, and

quasi experiments Nonrigorous inquiries somewhat resembling controlled experiments but lacking key elements such as pre- and posttesting and/or control groups.

time-series design A research design that involves measurements made over some period, such as the study of traffic-accident rates before and after lowering the speed limit.
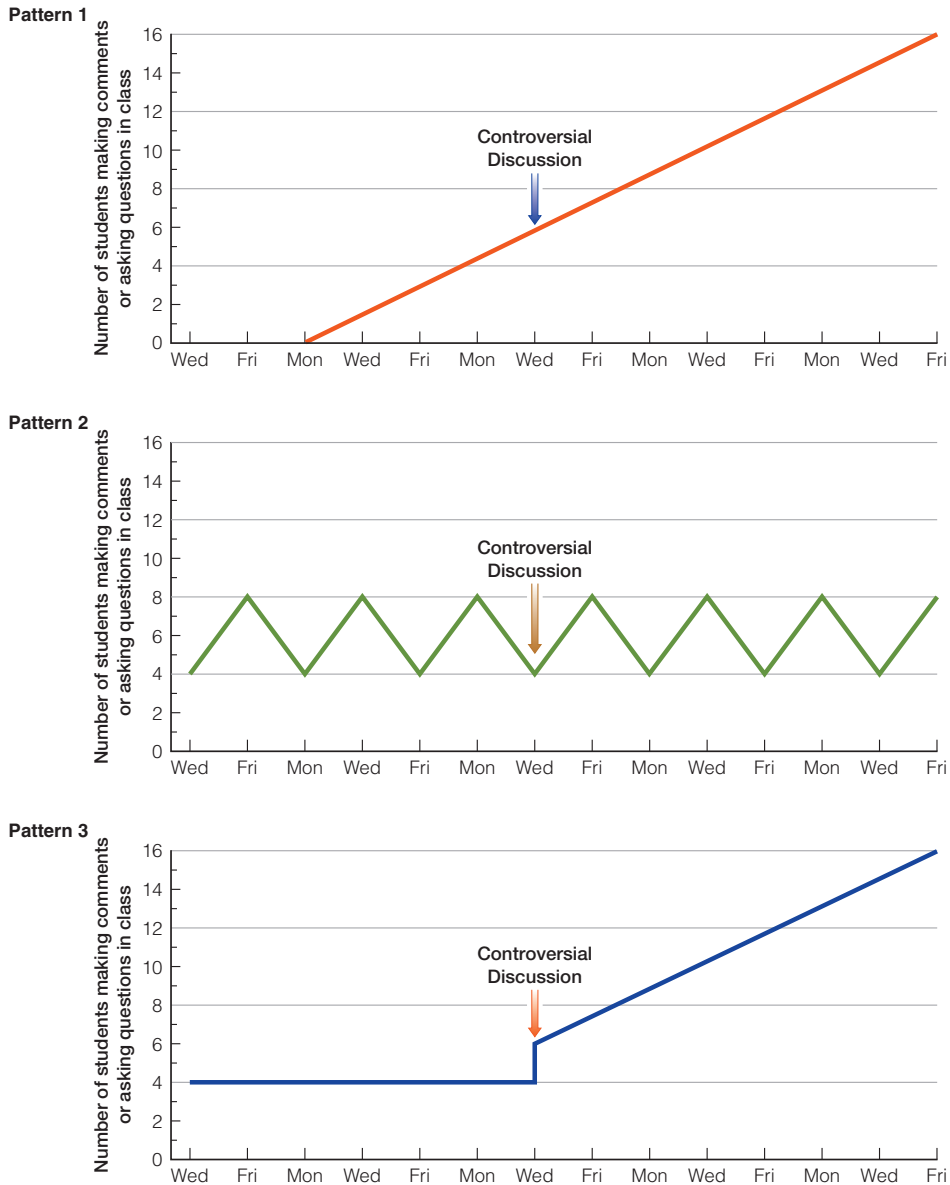
Pattern 1

Pattern 2

Pattern 3

FIGURE 12-2

Three Patterns of Class Participation in a Longer Historical Period

© Cengage Learning®

sometimes it decreases; the open discussion on that Wednesday simply came at a time when the level of participation was about to increase. More to the point, we note that class participation decreased again at the class following the alleged postdiscussion increase.

Only the third pattern in Figure 12-2 supports my contention that the open discussion mattered. As depicted there, the level of discussion before that Wednesday had been a steady four students

per class. Not only did the level of participation double following the day of the discussion, but it continued to increase afterward. Although these data do not protect us against the possible influence of some extraneous factor (I might also have mentioned that participation would figure into students' grades), they do exclude the possibility that the increase results from a process of maturation (indicated in the first pattern) or from regular fluctuations (indicated in the second).

Nonequivalent Control Groups

The time-series design just described involves only an "experimental" group; it doesn't provide the value to be gained from having a control group. Sometimes, when researchers can't create experimental and control groups by random assignment from a common pool, they can find an existing "control" group that appears similar to the experimental group. Such a group is called a **nonequivalent control group**. If an innovative foreign language program is being tried in one class in a large high school, for example, you may be able to find another foreign language class in the same school that has a very similar student population: one that has about the same composition in terms of grade in school, gender, ethnicity, IQ, and so forth. The second class, then, could provide a point of comparison even though it is not formally part of the study. At the end of the semester, you could give both classes the same foreign language test and then compare performances.

Here's how two junior high schools were selected for purposes of evaluating a program aimed at discouraging tobacco, alcohol, and drug use:

> The pairing of the two schools and their assignment to "experimental" and "control" conditions was not random. The local Lung Association had identified the school where we delivered the program as one in which administrators were seeking a solution to admitted problems of smoking, alcohol, and drug abuse. The "control" school was chosen as a convenient and nearby demographic match where administrators were willing to allow our surveying and breath-testing procedures. The principal of that school considered the existing program of health education to be effective and believed that the onset of smoking was relatively uncommon among his students. The communities served by the two schools were very similar. The rate of parental smoking reported by the students was just above 40 percent in both schools.
>
> *(McAlister et al. 1980: 720)*

In the initial set of observations, the experimental and control groups reported virtually the same (low) frequency of smoking. Over the 21 months of the study, smoking increased in both groups, but it increased less in the experimental group than in the control group, suggesting that the program affected students' behavior.

Multiple Time-Series Designs

Sometimes the evaluation of processes occurring outside of "pure" experimental controls can be made easier by the use of more than one time-series analysis. **Multiple time-series designs** are an improved version of the nonequivalent control group design just described. Carol Weiss (1972: 69) presents a useful example:

> An interesting example of multiple time series was the evaluation of the Connecticut crackdown on highway speeding. Evaluators collected reports of traffic fatalities for several periods before and after the new program went into effect. They found that fatalities went down after the crackdown, but since the series had had an unstable up-and-down pattern for many years, it was not certain that the drop was due to the program. They then compared the statistics with time-series data from four neighboring states where there had been no changes in traffic enforcement. Those states registered no equivalent drop in fatalities. The comparison lent credence to the conclusion that the crackdown had had some effect.

Although this study design is not as good as one in which subjects are assigned randomly, it's nonetheless an improvement over assessing the experimental group's performance without any comparison. That's what makes these designs quasi experiments instead of just fooling around. The key in assessing this aspect of evaluation studies is comparability, as the following example illustrates.

nonequivalent control group A control group that is similar to the experimental group but is not created by the random assignment of subjects. This sort of control group differs significantly from the experimental group in terms of the dependent variable or variables related to it.

multiple time-series designs The use of more than one set of data that were collected over time, as in accident rates over time in several states or cities, so that comparisons can be made.

Rural development, a growing concern in the poor countries of the world, has captured the attention and support of many rich countries. Through national foreign-assistance programs and international agencies such as the World Bank, the developed countries are in the process of sharing their technological knowledge and skills with the developing countries. Such programs have had mixed results, however. Often, modern techniques do not produce the intended results when applied in traditional societies.

Rajesh Tandon and L. Dave Brown (1981) undertook an experiment in which technological training would be accompanied by instruction in village organization. They felt it was important for poor farmers to learn how to organize and exert collective influence within their villages—getting needed action from government officials, for example. Only then would their new technological skills bear fruit.

Both intervention and evaluation were attached to an ongoing program in which 25 villages had been selected for technological training. Two poor farmers from each village had been trained in new agricultural technologies. Then they had been sent home to share their new knowledge with their village and to organize other farmers into "peer groups" who would assist in spreading that knowledge. Two years later, the authors randomly selected two of the 25 villages (subsequently called Group A and Group B) for special training and 11 other untrained groups as controls. A careful comparison of demographic characteristics showed the experimental and control groups to be strikingly similar, suggesting they were sufficiently comparable for the study.

The peer groups from the two experimental villages were brought together for special training in organization building. The participants were given some information about organizing and making demands on the government, and they were also given opportunities to act out dramas similar to the situations they faced at home. The training took three days.

The outcome variables considered by the evaluation all had to do with the extent to which members of the peer groups initiated group activities designed to improve their situation. Six types were studied. "Active initiative," for example, was defined as "active effort to influence persons or events affecting group members versus passive response or withdrawal" (Tandon and Brown 1981: 180). The data for evaluation came from the journals that the peer-group leaders had been keeping since their initial technological training. The researchers read through the journals and counted the number of initiatives taken by members of the peer groups. Two researchers coded the journals independently and compared their work to test the reliability of the coding process.

Figure 12-3 compares the number of active initiatives by members of the two experimental groups with those coming from the control groups. Similar results were found for the other outcome measures.

Notice two things about the graph. First, there is a dramatic difference in the number of initiatives by the two experimental groups as compared with the eleven controls. This would seem to confirm the effectiveness of the special training program. The orange line in the graph represents Group A, which received training in active initiatives, while the green line represents Group B, which also received the training. We see that the number of instances of active initiatives greatly increased for these two groups.

Second, notice that the number of initiatives also increased among the control groups. The researchers explain this latter pattern as a result of contagion. Because all the villages were near each other, the lessons learned by peer-group members in the experimental groups were communicated in part to members of the control villages. Although the control groups showed an increase in initiatives, the key evaluation here lies in comparing the experimental (A and B) groups with the controls.

This example illustrates the strengths of multiple time-series designs in situations where true experiments are inappropriate to the program being evaluated.

Qualitative Evaluations

Although I've laid out the steps involved in tightly structured, mostly quantitative evaluation research, evaluations can also be less structured and more qualitative. For example, Pauline Bart and Patricia O'Brien (1985) wanted to evaluate different ways to stop rape, so they undertook
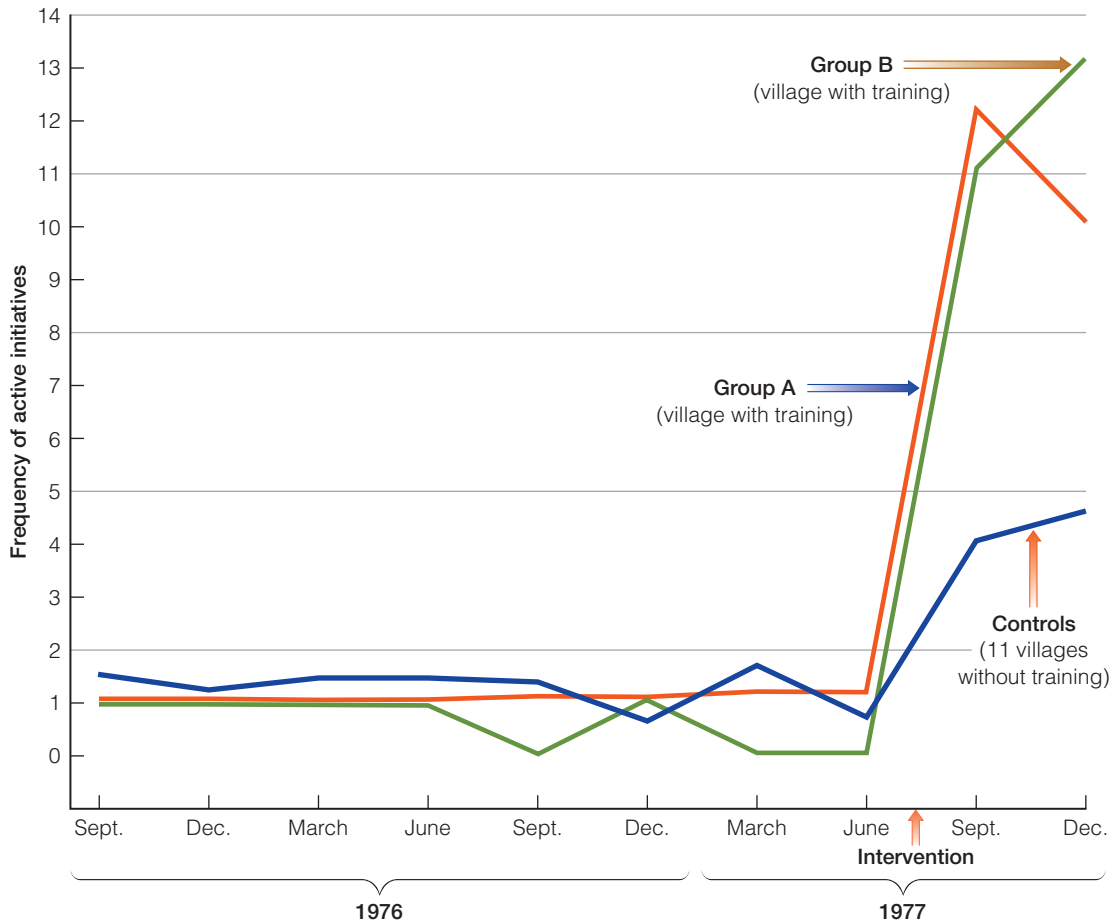
FIGURE 12-3

Active Initiatives over Time

Source: Rajesh Tandon and L. Dave Brown. 1981. "Organization-Building for Rural Development: An Experiment in India." *Journal of Applied Behavioral Science* (April–June): 182.

in-depth interviews with rape victims and with women who had successfully fended off rape attempts. As a general rule, they found that resistance (e.g., yelling, kicking, running away) was more likely to succeed than to make the situation worse, as women sometimes fear it will.

At times, even structured quantitative evaluations can yield unexpected qualitative results. Paul Steel is a social researcher specializing in the evaluation of programs aimed at pregnant drug users. One program he evaluated involved counseling by public-health nurses, who warned pregnant drug users that continued drug use would likely result in underweight babies whose skulls would be an average of 10 percent smaller than normal. In his in-depth interviews with program participants, however, he discovered

that the program omitted one important piece of information: that undersized babies were a bad thing. Many of the young women Steel interviewed thought that smaller babies would mean easier deliveries.

In another program, a local district attorney had instituted what would generally be regarded as a progressive, enlightened program. If a pregnant drug user were arrested, she could avoid prosecution if she would (1) agree to stop using drugs and (2) successfully complete a drug-rehabilitation program. Again, in-depth interviews suggested that the program did not always operate on the ground the way it did in principle. Specifically, Steel discovered that whenever a young woman was arrested for drug use, her fellow inmates would advise her to get pregnant

as soon as she was released on bail. That way, she would be able to avoid prosecution (personal communication, November 22, 1993).

The most effective evaluation research is one that combines qualitative and quantitative components. Making statistical comparisons is useful, and so is gaining an in-depth understanding of the processes producing the observed results—or preventing the expected results from appearing.

The evaluation of the Tanzanian soap opera, presented earlier in this chapter, employed several research techniques. I've already mentioned the listener surveys and data obtained from clinics. In addition, the researchers conducted numerous focus groups to probe more deeply into the impact the shows had on listeners. Also, content analyses were done on the soap opera episodes themselves and on the many letters received from listeners. Both quantitative and qualitative analyses were undertaken (Swalehe et al. 1995).

The soap opera research also offers an opportunity to see the impact of different cultures on the conduct of research. I had an opportunity to experience this firsthand when I consulted on the evaluation of soap operas being planned in Ethiopia. In contrast to the Western concern for confidentiality in social research, respondents selected for interviews in rural Ethiopian villages often took a special pride at being selected and wanted their answers broadly known in the community.

Or, sometimes, local researchers' desires to please the client got in the way of the evaluation. For example, some pilot episodes were tested in focus groups to determine whether listeners would recognize any of the social messages being communicated. The results were more encouraging than could have been expected. When I asked how the focus group subjects had been selected, the researcher described his introductory conversation: "We would like you to listen to some radio programs designed to encourage people to have small families, and we'd like you to tell us whether we've been successful." Not surprisingly, the small-family theme came through clearly to the focus group.

These experiences, along with earlier comments in previous sections, hint at the possibility of problems in the actual execution of evaluation research projects. Of course, all forms of research can run into problems, but evaluation research has a special propensity for it, as we shall now explore further.

The Social Context

This section looks at some of the logistical problems in evaluation research and presents some observations about using evaluation research results. The social context also raises special ethical issues; we'll discuss these at the end of the chapter.

Logistical Problems

In a military context, *logistics* refers to moving supplies around—making sure people have food, guns, and tent pegs when they need them. Here, I use it to refer to getting subjects to do what they're supposed to do, getting research instruments distributed and returned, and other seemingly simple tasks. These tasks are more challenging than you might guess!

Motivating Sailors

When Kent Crawford, Edmund Thomas, and Jeffrey Fink (1980) set out to find a way to motivate "low performers" in the U.S. Navy, they found out just how many problems can occur. The purpose of the research was to test a three-pronged program for motivating sailors who were chronically poor performers and often in trouble aboard ship. First, a workshop was to be held for supervisory personnel, training them in the effective leadership of low performers. Second, a few supervisors would be selected and trained as special counselors and role models—people the low performers could turn to for advice or just as sounding boards. Finally, the low performers themselves would participate in workshops aimed at training them to be more motivated and effective in their work and in their lives. The project was to be conducted aboard a particular ship, with a control group selected from sailors on four other ships.

To begin, the researchers reported that the supervisory personnel were not exactly thrilled with the program.

Not surprisingly, there was considerable resistance on the part of some supervisors toward dealing with these issues. In fact, their reluctance to assume ownership of the problem was reflected by "blaming" any of several factors that can contribute to their personnel problem. The recruiting system, recruit training, parents, and society at large were named as influencing low performance—factors that were well beyond the control of the supervisors.

(Crawford et al. 1980: 488)

Eventually, the reluctant supervisors came around and "this initial reluctance gave way to guarded optimism and later to enthusiasm" (1980: 489).

The low performers themselves were even more of a problem, however. The research design called for pre- and posttesting of attitudes and personalities, so that changes brought about by the program could be measured and evaluated.

Unfortunately, all of the LPs (Low Performers) were strongly opposed to taking these so-called personality tests and it was therefore concluded that the data collected under these circumstances would be of questionable validity. Ethical concerns also dictated that we not force "testing" on the LPs.

(Crawford et al. 1980: 490)

As a consequence, the researchers had to rely on interviews with the low performers and on the judgments of supervisors for their measures of attitude change. The subjects continued to present problems, however.

Initially, the ship's command ordered 15 low performers to participate in the experiment. Of the 15, however, one went into the hospital, another was assigned duties that prevented participation, and a third went "over the hill" (absent without leave). Thus, the experiment began with 12 subjects. But before it was completed, three more subjects completed their tour of duty and left the Navy, and another was thrown out for disciplinary reasons. The experiment concluded, then, with 8 subjects. Although the evaluation pointed to positive results, the very small number of subjects warranted caution in any generalizations from the experiment.

The special, logistical problems of evaluation research grow out of the fact that it occurs within the context of real life. Although evaluation research is modeled after the experiment—which suggests that the researchers have control over what happens—it takes place within frequently uncontrollable daily life. Of course, the participant-observer in field research doesn't have control over what is observed either, but that method doesn't strive for control. Given the objectives of evaluation research, lack of control can create real dilemmas for the researcher.

Administrative Control

As suggested in the previous example, the logistical details of an evaluation project often fall to program administrators. Let's suppose you're evaluating the effects of a "conjugal visit" program on the morale of married prisoners. The program allows inmates periodic visits from their spouses during which they can have sexual relations. On the fourth day of the program, a male prisoner dresses up in his wife's clothes and escapes. Although you might be tempted to assume that his morale was greatly improved by escaping, that turn of events would complicate your study design in many ways. Perhaps the warden will terminate the program altogether, and where's your evaluation then? Or, if the warden is brave, he or she may review the files of all those prisoners you selected randomly for the experimental group and veto the "bad risks." There goes the comparability of your experimental and control groups. As an alternative, stricter security measures may be introduced to prevent further escapes, but the security measures may have a dampening effect on morale. So the experimental stimulus has changed in the middle of your research project. Some of the data will reflect the original stimulus; other data will reflect the modification. Although you'll probably be able to sort it all out, your carefully designed study has become a logistical snake pit.

Or suppose you've been engaged to evaluate the effect of race-relations lectures on prejudice in the army. You've carefully studied the soldiers available to you, and you've randomly assigned some to attend the lectures and others to stay away. The rosters have been circulated weeks in advance, and at the appointed day and hour, the lectures begin. Everything seems to be going smoothly until you begin processing the files: The names don't match. Checking around, you

Research in Real Life

Testing Soap Operas in Tanzania

William N. Ryerson

President and Founder, Population Media Center

Twende na Wakati (*Let's Go with the Times*) has been broadcast on Radio Tanzania since mid-1993 with support from the United Nations Population Fund. The program was designed to encourage family-planning use and AIDS prevention measures.

There were many different elements to the research. One was a nationwide, random-sample survey given prior to the first airing of the soap opera in June 1993 and then annually after that. Many interviewers faced particularly interesting challenges. For example, one interviewer, Fridolan Banzi, had never been in or on water in his life and couldn't swim. He arranged for a small boat to take him through the rough waters of Lake Victoria so he could carry out his interviews at a village

that had no access by road. He repeated this nerve-wracking trip each year afterward in order to measure the change in that village.

Another interviewer, Mr. Tende, was invited to participate in a village feast that the villagers held to welcome him and to indicate their enthusiasm about having been selected for the study. They served him barbequed rats. Though they weren't part of his normal diet, he ate them anyway to be polite and to ensure that the research interviews could be carried out in that village.

Still another interviewer, Mrs. Masanja, was working in a village in the Pwani region along the coast of the Indian Ocean when cholera broke out in that village. She wisely chose to abandon the interviews there, which reduced the 1993 sample size by one ward. The unsung heroes of this research, the Tanzanian interviewers, deserve a great deal of credit for carrying out this important work under difficult circumstances.

---

discover that military field exercises, KP duty, and a variety of emergencies required some of the experimental subjects to be elsewhere at the time of the lectures. That's bad enough, but then you learn that helpful commanding officers sent others to fill in for the missing soldiers. And whom do you suppose they picked to fill in? Soldiers who didn't have anything else to do or who couldn't be trusted to do anything important. You might learn this bit of information a week or so before the deadline for submitting your final report on the impact of the race-relations lectures.

These are some of the logistical problems confronting evaluation researchers. You need to be familiar with the problems to understand why some research procedures may not measure up to the design of the classical experiment. As you read reports of evaluation research, however, you'll find that—my earlier comments notwithstanding—it is possible to carry out controlled social research in conjunction with real-life experiments.

The Research in Real Life box, "Testing Soap Operas in Tanzania," describes some of the logistical problems involved in the research discussed at the outset of this chapter.

Use of Research Results

One more facts-of-life aspect of evaluation research concerns how evaluations are used.

Because the purpose of evaluation research is to determine the success or failure of social interventions, you might think it reasonable that a program would automatically be continued or terminated based on the results of the research.

Reality isn't that simple and reasonable, however. Other factors intrude on the assessment of evaluation research results, sometimes blatantly and sometimes subtly. Undoubtedly every evaluation researcher can point to studies he or she conducted—studies providing clear research results and obvious policy implications—that were ignored.

The 1990s saw the passage of "three strikes" laws at the federal level and in numerous states. The intention was to reduce crime rates by locking up "career criminals." Under the 1994 California law, for example, having a past felony conviction would double your punishment when you were convicted of your second felony, and the third felony conviction would bring a mandatory sentence of 25 years to life. Over the years, only California has enforced such laws with any vigor.

Those who supported the passage of "three strikes" legislation, such as Bill Jones, quoted earlier, were quick to link the dramatic drop in crime rates during the 1990s to the new policy of getting tough with career criminals. While acknowledging that "three strikes" may not be the only cause of the drop in crime, Jones added,

"If you can have a 51 percent reduction in the homicide rate in five years, I would guarantee you three strikes is a big part of the reason."

In spite of the politicians' guarantees, other observers have looked for additional evidence to support the impact of "three strikes" laws. Some critics of these laws, for example, have noted that crime rates have been dropping dramatically across the country, not only in California but in states that have no "three strikes" laws and in those where the courts have not enforced the "three strikes" laws that exist. In fact, crime rates have dropped in those California counties that have tended to ignore that state's law. Moreover, the drop in California crime rates began before the "three strikes" law went into effect.

In 1994, Peter Greenwood and his colleagues at the Rand Corporation estimated that implementation of the law would cost California's criminal justice system approximately $5.5 billion more per year, especially in prison costs as "career criminals" were sentenced to longer terms. Although the Rand group did not deny that the "three strikes" legislation would have some impact on crime—those serving long terms in prison can't commit crimes on the streets—a follow-up study (Greenwood, Rydell, and Model 1996) suggested it was an inefficient way of attacking crime. They estimated that a million dollars spent on "three strikes" would prevent 60 crimes, whereas the same amount spent on programs encouraging high school students to stay in school and graduate would prevent 258 crimes.

Criminologists have long recognized that most crimes are committed by young men. Focusing attention on older "career criminals" has little or no affect on the youthful offenders. In fact, "three strikes" sentences disproportionately fall on those approaching the end of their criminal careers by virtue of growing older.

In a more general critique, John Irwin and James Austin (1997) suggest that people in the United States tend to overuse prisons as a solution to crime, ignoring other, more effective, solutions. Often, imprisonment causes problems more serious than those it was intended to remedy.

As with many other social interventions, however, much of the support for "three strikes" laws in California and elsewhere stems mostly from public emotions about crime and the political implications of such emotions. Thus,

evaluation research on these laws may eventually bring about changes, but its impact is likely to be much slower than you might logically expect.

There are three important reasons why the implications of the evaluation research results are not always put into practice. First, the implications may not always be presented in a way that the nonresearchers can understand. Second, evaluation results sometimes contradict deeply held beliefs. People thought Copernicus was crazy when he said the earth revolved around the sun. Anybody could tell the earth was standing still. The third barrier to the use of evaluation results is vested interests. If I've devised a new rehabilitation program that I'm convinced will keep ex-convicts from returning to prison, and if people have taken to calling it "The Babbie Plan," how do you think I'm going to feel when your evaluation suggests the program doesn't work? I might apologize for misleading people, fold up my tent, and go into another line of work. More likely, I'd call your research worthless and begin intense lobbying with the appropriate authorities to have my program continue.

In the earlier example of the evaluation of drivers' education, Philip Hilts reported some of the reactions to the researchers' preliminary results:

> Ray Burneson, traffic safety specialist with the National Safety Council, criticized the study, saying that it was a product of a group (NHTSA) run by people who believe "that you can't do anything to train drivers. You can only improve medical facilities and build stronger cars for when the accidents happen. . . . This knocks the whole philosophy of education."
>
> *(1981: 4)*

By its nature, evaluation research takes place in the midst of real life, affecting it and being affected by it. Here's another example, well known to social researchers.

Rape Reform Legislation

For years, many social researchers and other observers have noted certain problems with the prosecution of rape cases. All too often, it is felt, the victim's time on the witness stand severely prolongs the suffering. Frequently the defense lawyers portray the rape victim as having

encouraged the sex act and being of shady moral character; other personal attacks are intended to deflect responsibility from the accused rapist.

Criticisms such as these have resulted in a variety of state-level legislation aimed at remedying the problems. Cassie Spohn and Julie Horney (1990) were interested in tracking the impact of such legislation. The researchers summarize the ways in which new laws were intended to make a difference:

> The most changes are: (1) redefining rape and replacing the single crime of rape with a series of graded offenses defined by the presence or absence of aggravating conditions; (2) changing the consent standard by eliminating the requirement that the victim physically resist her attacker; (3) eliminating the requirement that the victim's testimony be corroborated; and (4) placing restrictions on the introduction of evidence of the victim's prior sexual conduct.
>
> *(1990: 2)*

It was generally expected that such legislation would encourage more women to file a police report when they've been raped, which would thus increase convictions when the cases are brought to court. To examine the latter expectation, the researchers focused on the period from 1970 to 1985 in Cook County, Illinois: "Our data file includes 4,628 rape cases, 405 deviate sexual assault cases, 745 aggravated criminal sexual assault cases, and 37 criminal sexual assault cases" (1990: 4). Table 12-1 shows some of what they discovered.

Spohn and Horney summarized these findings as follows:

> The only significant effects revealed by our analyses were increases in the average maximum prison sentences; there was an increase of almost 48 months for rape and of almost 36 months for sex offenses. Because plots of the data indicated an increase in the average sentence before the reform took effect, we modeled the series with the intervention moved back one year earlier than the actual reform date. The size of the effect was even larger and still significant, indicating that the effect should not be attributed to the legal reform.
>
> *(1990: 10)*

TABLE 12-1

Analysis of Rape Cases Before and After Legislation

	Rape	
	Before (N = 2,252)	After (N = 2,369)
Outcome of case		
Convicted of original charge	45.8%	45.4%
Convicted of another charge	20.6	19.4
Not convicted	33.6	35.1
Median prison sentence in months		
For those convicted of original charge	96.0	144.0
For those convicted of another charge	36.0	36.0

© Cengage Learning®

Notice in the table that there was virtually no change in the percentages of cases ending in conviction for rape or some other charge (e.g., assault). Hence the change in laws didn't have any effect on the likelihood of conviction. As the researchers note, the one change that *is* evident—an increase in the length of sentences—cannot be attributed to the reform legislation itself.

In addition to the analysis of existing statistics, Spohn and Horney interviewed judges and lawyers to determine what they felt about the impact of the laws. Their responses were somewhat more encouraging.

> Judges, prosecutors and defense attorneys in Chicago stressed that rape cases are taken more seriously and rape victims treated more humanely as a result of the legal changes. These educative effects clearly are important and should please advocates of rape reform legislation.
>
> *(1990: 17)*

Thus, the study found other effects besides the qualitative results the researchers looked for. This study demonstrates the importance of following up on social interventions to determine whether, in what ways, and to what degree they accomplished their intended results.

Preventing Domestic Violence

In a somewhat similar study, researchers in Indianapolis focused their attention on the

problem of wife battering, with a special concern for whether prosecuting the batterers can lead to subsequent violence. David Ford and Mary Jean Regoli (1992) set about studying the consequences of various options for prosecution allowed within the "Indianapolis Prosecution Experiment" (IPE).

Wife-battering cases can follow a variety of patterns, as Ford and Regoli summarize:

> After a violent attack on a woman, someone may or may not call the police to the scene. If the police are at the scene, they are expected to investigate for evidence to support probable cause for a warrantless arrest. If it exists, they may arrest at their discretion. Upon making such an on-scene arrest, officers fill out a probable cause affidavit and slate the suspect into court for an initial hearing. When the police are not called, or if they are called but do not arrest, a victim may initiate charges on her own by going to the prosecutor's office and swearing out a probable cause affidavit with her allegation against the man. Following a judge's approval, the alleged batterer may either be summoned to court or be arrested on a warrant and taken to court for his initial hearing.
>
> *(1992: 184)*

What if a wife brings charges against her husband and then reconsiders later on? Many courts have a policy of prohibiting such actions, in the belief that they are serving the interests of the victim by forcing the case to be pursued to completion. In the IPE, however, some victims are offered the possibility of dropping the charges if they so choose later in the process. In addition, the court offers several other options. Because wife battering is largely a function of sexism, stress, and an inability to deal with anger, some of the innovative possibilities in the IPE involve educational classes with anger-control counseling.

If the defendant admits his guilt and is willing to participate in an anger-control counseling program, the judge may postpone the trial for that purpose and can later dismiss the charges if the defendant successfully completes the program. Alternatively, the defendant may be tried and, if found guilty, be granted probation provided he participates in the anger-control

program. Finally, the defendant can be tried and, if found guilty, be given a conventional punishment such as imprisonment.

Which of these possibilities most effectively prevents subsequent wife battering? That's the question Ford and Regoli addressed. Here are some of their findings.

First, their research shows that men who are brought to court for a hearing are less likely to continue beating their wives, no matter what the outcome of the hearing. Simply being brought into the criminal justice system has an impact.

Second, women who have the right to drop charges later on are less likely to be abused subsequently than those who do not have that right. In particular, the combined policies of arresting defendants by warrant and allowing victims to drop charges provides victims with greater security from subsequent violence than any of the other prosecution policies do.

However, giving victims the right to drop charges has a somewhat strange impact. Women who exercise that right are more likely to be abused later than those who insist on the prosecution proceeding to completion. The researchers interpret this as showing that future violence can be decreased when victims have a sense of control supported by a clear and consistent alliance with criminal justice agencies.

> A decisive system response to any violation of conditions for pretrial release, including of course new violence, should serve notice that the victim-system alliance is strong. It tells the defendant that the victim is serious in her resolve to end the violence and that the system is unwavering in its support of her interest in securing protection.
>
> *(Ford and Regoli 1992: 204)*

The effectiveness of anger-control counseling cannot be assessed simply. Policies aimed at getting defendants into anger-control counseling seem to be relatively ineffective in preventing new violence. The researchers noted, however, that the policy effects should not be confused with actual counseling outcomes. Some defendants scheduled for treatment never received it. Considerably more information on implementing counseling is needed for a proper evaluation.

Moreover, the researchers cautioned that their results point to general patterns, and that

battered wives must choose courses of action appropriate to their particular situations and should not act blindly on the basis of the overall patterns. The research is probably more useful in what it says about ways of structuring the criminal justice system (giving victims the right to drop charges, for example) than in guiding the actions of individual victims.

Finally, the IPE offers an example of a common problem in evaluation research. Often, actual practices differ from what might be expected in principle. For example, the researchers considered the impact of different alternatives for bringing suspects into court: Specifically, the court can issue either a summons ordering the husband to appear in court or a warrant to have the husband arrested. The researchers were concerned that having the husband arrested might actually add to his anger over the situation. They were somewhat puzzled, therefore, to find no difference in the anger of husbands summoned or arrested.

The solution of the puzzle lay in the discrepancy between principle and practice:

> Although a warrant arrest should in principle be at least as punishing as on-scene arrest, in practice it may differ little from a summons. A man usually knows about a warrant for his arrest and often elects to turn himself in at his convenience, or he is contacted by the warrant service agency and invited to turn himself in. Thus, he may not experience the obvious punishment of, say, being arrested, handcuffed, and taken away from a workplace.
>
> *(Ford 1989: 9–10)*

In summary, many factors besides the scientific quality of evaluation research affect how its results are used. And, as we saw earlier, factors outside the evaluator's control can affect the quality of the study itself. But this "messiness" is balanced by the potential contributions that evaluation research can make toward the betterment of human life.

The Sabido Methodology

One of the clearest illustrations of the uses of evaluation research results can be found in the omnibus methodology developed by Miguel Sabido for the use of "Entertainment-Education" (E-E) projects to promote social programs. The example of *Twende na Wakati* at the outset of this chapter illustrated the methods initially developed by Sabido in the 1970s when he was Vice President for Research in Mexico's national broadcasting company Televisa. Sabido's first projects used television novellas to promote literacy and family planning. They were so successful that those methods have been used to promote a variety of social issues in the subsequent decades.

In part, the Sabido methodology concerns the nature of the radio or television dramas: particularly the kinds of characters portrayed. Some characters represent traditional points of view, some represent the modern views that the programming is designed to promote, and some represent a "transitional" point of view—they begin with traditional views but eventually shift to the modern views. Typically, when a transitional character signs up for literacy classes, thousands of audience members do the same shortly thereafter. When the transitional character begins using condoms for family planning or safe sex, family planning clinics are mobbed the next day by men wanting condoms.

The Sabido methodology extends beyond character definitions and plot structures. An E-E project begins with thorough research into the society where the change is being planned. A project in Ethiopia by the Population Media Center, for example, aimed to lower the birthrate, encourage safe-sex practices, and enhance the status of women. The production of radio serial dramas was preceded by extensive research into the existing situations regarding the project's aims. What was the birthrate? How did it differ in different regions of the country and among different ethnic groups? What were the attitudes toward family planning? In part these questions were answered through national surveys. At the same time, qualitative researchers went into the countryside to observe rural villages, talking with residents and sometimes recording the sounds of village life.

This formative research provided the writers with ideas about issues to be raised and how to raise them. For example, the research indicated that in some regions, abduction was still a common method of mate selection: A man would kidnap a young woman, sexually assaulting her and holding her prisoner until she would

Research in Real Life

Chinese Public Opinion

One of the consequences of the 1949 revolution in China was the cancellation of sociology as a field of study in Chinese universities. Some Chinese sociologists, like the respected Fei Xiao Tung, continued to conduct social research, monitoring the quality of life among citizens, especially those living in rural villages. Trained by the renowned Polish-British anthropologist, Bronislaw Malinowski, Fei primarily used qualitative, ethnographic methods. Whenever Chairman Mao would invite criticism of his regime, Professor Fei would be one of the first in line to report his research on what wasn't working and how to improve life in China. His findings were not always well received by the government, but he was persistent.

Finally, in 1979, the Peoples Republic of China reinstated sociology as a field of study, and Professor Fei was given the responsibility of creating the institutions necessary for the emergence of new generations of Chinese social scientists. Today, Chinese scholars are actively examining all aspects of social life. For example, the Canton Public Opinion Research (C-POR) Center in Guangzhou seeks to uncover problems in the quality of life of both urban and rural people, and draw the government's attention to those problems.

You can explore this research at the center's website: www.c-por.org/. As you will discover, most of the content is in Chinese characters, but a short detour to Google Translator, www.google.com/language_tools, will solve your problem. Here are only a few report titles from late 2011:

"Guangzhou City Public Health Assessment Survey";
"Guangzhou Environmental Protection Status Follow-Up Survey of Public Evaluation";
"Guangzhou Food Safety Assessment Survey of Public Opinion";
"Guangzhou Government 'Window Service' Satisfaction Follow-Up Investigation";
"Evaluation of Public Health in Rural Village";
"Rural Education, Supervision, Valuation of the Low Mountains of Northern Guangdong Villagers";
"Education and Supervision by the Villagers Recognized Quality Control to Be Improved";
"Beijing People's Life Experience Is Best."

When I had the opportunity to meet with some of the C-POR researchers in Shanghai in 2010, I was struck by their enthusiasm for using social research methods to improve the lot of the Chinese people. They are worthy successors to Professor Fei Xiao Tung.

---

consent to be his wife. The formative research also revealed a widespread belief that condoms were infected with HIV, thus meaning that condom use increased the risk of AIDS rather than reducing it.

The initial research also provided a baseline for subsequent evaluations. By knowing public opinion toward family planning prior to the radio programs, researchers could determine how much these opinions had changed afterward. Preprogramming measures of the use of family planning centers could be compared with use levels afterward. Many of these evaluation efforts ran concurrently with the radio programming. For example, regular focus groups were used to monitor public reactions to each of the serial installments, examining whether people were reacting as intended.

The Sabido methodology provides an excellent illustration of how research methods can be used to construct and evaluate social action programs aimed at resolving social problems.

As you can see, evaluation research can provide a unique and powerful tool for effecting social change. However, it can also be useful on a personal level, in everyday situations, for such purposes as improving your grades, losing weight, making friends, and influencing people. Moreover, evaluation research is hardly limited to the United States, as the Research in Real Life box, "Chinese Public Opinion" illustrates.

Social Indicators Research

Let's now look at a type of research that combines what you've learned about evaluation research and about the analysis of existing data. A rapidly growing field in social research involves the development and monitoring of **social indicators**, aggregated statistics that

social indicators Measurements that reflect the quality or nature of social life, such as crime rates, infant mortality rates, number of physicians per 100,000 population, and so forth. Social indicators are often monitored to determine the nature of social change in a society.

reflect the social condition of a society or social subgroup. Researchers use social indicators to monitor aspects of social life in much the way that economists use indexes such as gross national product (GNP) per capita as an indicator of a nation's economic development.

Suppose we wanted to compare the relative health conditions in different societies. One strategy would be to compare their death rates (number of deaths per 1,000 population). Or, more specifically, we could look at infant mortality: the number of infants who die during their first year of life among every 1,000 births. Depending on the particular aspect of health conditions we were interested in, we could devise any number of other measures: physicians per capita, hospital beds per capita, days of hospitalization per capita, and so forth. Notice that intersocietal comparisons are facilitated by calculating per capita rates (dividing by the size of the population).

Before we go further, recall from Chapter 11 the problems involved in using existing statistics. In a word, they're sometimes unreliable, reflecting their modes of collection, storage, and calculation. This is not to invalidate this important resource but to remind us that we must be sure they measure what we wish to study or at least recognize how they differ. With this in mind, we'll look at some of the ways we can use social indicators for evaluation research on a large scale.

The Death Penalty and Deterrence

Does the death penalty deter capital crimes such as murder? This question is hotly debated every time a state considers eliminating or reinstating capital punishment and every time someone is executed. Those supporting capital punishment often argue that the threat of execution will deter potential murderers from killing people. Opponents of capital punishment often argue that it has no effect in that regard. Social indicators can help shed some light on the question.

If capital punishment actually deters people from committing murder, then we should expect to find murder rates lower in those states that have the death penalty than in those that do not. The relevant comparisons in this instance are not only possible, they've been compiled and published. William Bailey (1975) compiled

data that directly contradicted the view that the death penalty deters murderers. In both 1967 and 1968, those states with capital punishment had dramatically higher murder rates than those without capital punishment. Some people criticized the interpretation of Bailey's data, saying that most states with capital punishment on the books had not actually executed any prisoner in recent years, which could thus explain why the death penalty hadn't seemed to work as a deterrent. Further analysis, however, contradicts this explanation. When Bailey compared those states that hadn't used the death penalty with those that had, he found no real difference in murder rates.

Another counter explanation is possible, however. It could be the case that the interpretation given Bailey's data was *backward.* Maybe the existence of the death penalty as an option was a consequence of high murder rates: Those states with high rates instituted it; those with low rates didn't institute it or repealed it if they had it on the books. It could be the case, then, that instituting the death penalty would bring murder rates down. Not so, however. Analyses over time do not show an increase in murder rates when a state repeals the death penalty nor a decrease in murders when one is instituted. A more recent examination by Bailey and Ruth Peterson (1994) confirmed the earlier findings and also indicated that law enforcement official doubted the deterrent effect. Further, the pattern observed by Bailey in 1967 and 1968 has persisted over time, even when we take into account the substantial increase in the overall murder rate. In 2006, for example, the 38 death-penalty states had a combined murder rate of 5.90 per 100,000, compared with a combined murder rate of 3.85 among the 12 states that lack the death penalty (U.S. Bureau of the Census 2009: 17, 189).

Notice from the preceding discussion that it's possible to use social indicators data for comparison across groups either at one time or across some period of time. Often, doing both sheds the most light on the subject.

At present, work on the use of social indicators is proceeding on two fronts. On the one hand, researchers are developing increasingly refined indicators—finding which indicators of a general variable are the most useful in monitoring social life. At the same time, research is being

devoted to discovering the relationships among variables within whole societies.

As with many aspects of social research, the Internet has become a valuable resource. To pursue the possibilities of social indicators, you might check out Sociometrics Corporation, for example. Or simply search for "social indicators" using one of the web search engines.

Computer Simulation

An exciting prospect for social indicators research lies in the area of computer simulation. As researchers begin compiling mathematical equations describing the relationships that link social variables to one another (for example, the relationship between growth in population and the number of automobiles), those equations can be stored and linked to one another in the computer. With a sufficient number of adequately accurate equations on tap, researchers one day will be able to test the implications of specific social changes by computer rather than in real life.

Suppose a state contemplated doubling the size of its tourism industry, for example. We could enter that proposal into a computer-simulation model and receive a description of all the direct and indirect consequences of the increase in tourism. We could know what new public facilities would be required, which public agencies such as police and fire departments would have to be increased and by how much, what the labor force would look like, what kind of training would be required to provide it, how much new income and tax revenue would be produced, and so forth, through all the intended and unintended consequences of the action. Depending on the results, the public planners might say, "Suppose we increased the industry only by half," and have a new printout of consequences immediately.

An early illustration of computer simulation linking social and physical variables can be found in the research of Donella and Dennis Meadows and their colleagues at Dartmouth and the Massachusetts Institute of Technology (Meadows et al. 1972, 1992). They took as input data known and estimated reserves of various nonreplaceable natural resources (for example, oil, coal, iron), past patterns of population and economic growth, and the relationships between growth

and use of resources. Using a complex simulation model, they were able to project, among other things, the probable number of years that various resources would last in the face of alternative usage patterns in the future. Going beyond the initially gloomy projections, such models also make it possible to construct less-gloomy futures, specifying the actions required to achieve them. Clearly, the value of computer simulation is not limited to evaluation research, though it can serve an important function in that regard.

This potentiality points to the special value of evaluation research in general. Throughout human history, we've been tinkering with our social arrangements, seeking better results. Evaluation research provides a means for us to learn right away whether a particular tinkering really makes things better. Social indicators allow us to make that determination on a broad scale; coupling them with computer simulation opens up the possibility of knowing how much we would like a particular intervention, without having to experience its risks.

Ethics and Evaluation Research

As we have seen, evaluation research is by nature interwoven with real-world issues. Sometimes the social interventions being evaluated raise ethical issues. Evaluating the impact of busing school children to achieve educational integration will throw the researchers directly into the political, ideological, and ethical issues of busing itself. It's not possible to evaluate a sex-education program in elementary schools without becoming involved in the heated issues surrounding sex education itself, and the researcher will find remaining impartial difficult. The evaluation study design will *require* that some children receive sex education—in fact, you may very well be the one who decides which children do. (From a scientific standpoint, you *should* be in charge of selection.) This means that when parents become outraged that their child is being taught about sex, you'll be directly responsible.

Now let's look on the bright side. Maybe the experimental program is of great value to those participating in it. Let's say that the new industrial safety program being evaluated reduces

injuries dramatically. What about the control-group members who were deprived of the program by the research design? The evaluators' actions could be an important part of the reason that a control-group subject suffered an injury.

Sometimes the name of evaluation research has actually served as a mask for unethical behavior. In Chapter 9 I discussed push polls, which pretend to evaluate the impact of various political campaign accusations but intend to spread malicious misinformation. That's not the worst example, however, as you'll recall from the discussion of the Tuskegee experiments, in Chapter 3.

Even in the most legitimate evaluation research, the researcher almost always faces pressure from the people affected by the evaluation. Often, as in the case of pharmaceutical testing, for example, those paying for the research may want a particular result. Evaluation researchers, therefore, often find themselves under pressure to produce a particular finding.

I'm sure there's no need to point out that researchers must not be swayed by personal desires or sponsors' demands in the design, execution, and analysis of results; manipulating research to produce a desired result is never acceptable. This is particularly important in the case of evaluation research, in that the real-world setting can create serious and far-reaching consequences for the people involved. Imagine a medical researcher slanting drug-testing results to suggest a new drug is more effective than it is or covering up the negative side effects of the drug, so that the drug is given to patients who will not benefit from it or will actually be harmed by the "unknown" side effects. Or imagine that an evaluation of a prison rehabilitation program has been slanted to make the program seem more effective than it is. Limited resources might be diverted to support the ineffective program and possibly even harm the prisoners subjected to it.

My purpose in these comments has not been to cast a shadow on evaluation research. Rather, I want to bring home the real-life consequences of the evaluation researcher's actions. Ultimately, all social research has ethical components.

I will close this discussion with a somewhat different observation made by Donald T. Campbell in 1976. In what has come to be known as Campbell's law, he observed, "The

more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor" (54). One example of this is what educators refer to as "teaching to the test." If teachers are to be evaluated on the basis of how well their students perform on a standardized test, instruction tends to focus on that test rather than on the subject matter more generally. Similarly, when those managing stock portfolios are compensated on the basis of how many stocks have been traded, there is a temptation to trade stocks that might more wisely be held. Or, when police departments are judged as to their ability to lower assault rates in the city, there will be a temptation to categorize and report incidents as lesser offenses.

Thus, we see that evaluation research is sometimes a part of the process it seeks to evaluate and that it can have unintended consequences. This is another example of the recursive nature of social research, discussed in Chapter 1.

MAIN POINTS

Introduction

- Evaluation research is a form of applied research that studies the effects of social interventions.

Topics Appropriate for Evaluation Research

- Topics appropriate for evaluation research include needs assessment studies, cost-benefit studies, monitoring studies, and program evaluations/outcome assessments.
- Evaluation research is sometimes coupled with the intentions of participatory action research.

Formulating the Problem: Issues of Measurement

- A careful formulation of the problem, including relevant measurements and criteria of success or failure, is essential in evaluation research. In particular, evaluators must carefully specify outcomes, measure experimental contexts, specify the intervention being studied and the population targeted by the intervention, and decide whether to use existing measures or devise new ones.

Types of Evaluation Research Designs

- Evaluation researchers typically use experimental or quasi-experimental designs. Examples of quasi-experimental designs include time-series studies and the use of nonequivalent control groups.

- Evaluators can also use qualitative methods of data collection. Both quantitative and qualitative data analyses can be appropriate in evaluation research, sometimes in the same study.

The Social Context

- Evaluation research entails special logistical problems because it's embedded in the day-to-day events of real life.
- The implications of evaluation research won't necessarily be put into practice, especially if they conflict with official points of view.

Social Indicators Research

- Social indicators can provide an understanding of broad social processes.
- Computer-simulation models hold the promise of allowing researchers to study the possible results of social interventions without having to incur those results in real life.

Ethics and Evaluation Research

- Sometimes the social interventions being assessed in evaluation research themselves raise ethical issues.
- Evaluation research may entail added pressure to produce specific results, as desired by interested parties.
- Fraudulent research results in an evaluation study can have severer consequences than consequences produced by other types of research.

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.

cost-benefit studies

evaluation research

monitoring studies

multiple time-series designs

needs assessment studies

nonequivalent control group

program evaluation/ outcome assessment

quasi experiments

social indicators

time-series design

PROPOSING SOCIAL RESEARCH: EVALUATION RESEARCH

Evaluation research represents a research purpose rather than a particular method. In the proposal, you need to spell out the type of evaluation you're conducting and perhaps the implications of various possible outcomes.

In earlier assignments, you'll have spelled out the data-collection and measurement methods to be used in your study. If your study is designed to determine the success or failure of a program, you may also want to specify the research results that will be deemed a positive or negative assessment in that regard. This may not always be appropriate or possible, but it adds integrity to the evaluation process when it can be done.

REVIEW QUESTIONS AND EXERCISES

1. Suppose a community establishes an alcohol- and drug-free teen center as a way of reducing the use of alcohol and drugs by teenagers. Describe how you might go about evaluating the effectiveness of the center. Indicate whether your design would be experimental, quasi-experimental, or qualitative (or some combination of these).

2. Review the evaluation of the Navy low-performer program discussed in the chapter. Redesign the program and the evaluation to handle the problems that appeared in the actual study.

3. Discuss some of the potential political and ethical issues that might be involved in the study you described in Exercise 1.

4. Take a minute to think of the many ways your society has changed during your own lifetime. Specify three or four social indicators that could be used in monitoring the effects of at least one of those changes on the quality of life in your society.

5. The U.S. Bureau of Prisons engages in evaluation research regarding various aspects of prison operations. Locate one of their studies on the web and write a short summary of the study design and the findings.

In this part of the book, we'll discuss the analysis of social research data, and we'll examine the steps that separate observation from the final reporting of findings.

In Chapter 1, I made a fundamental distinction between qualitative and quantitative data. In the subsequent discussions, we've seen that many of the fundamental concerns in social research apply equally to both types of data. The analysis of qualitative and quantitative data, however, are quite different and will be discussed separately.

Before outlining the specifics of Part 4, I want to offer an observation about the ease or difficulty of producing high-quality data analyses, as represented in the following table, where "1" is the easiest to do and "4" is the hardest.

	Simplistic	Sophisticated
Qualitative	1	4
Quantitative	2	3