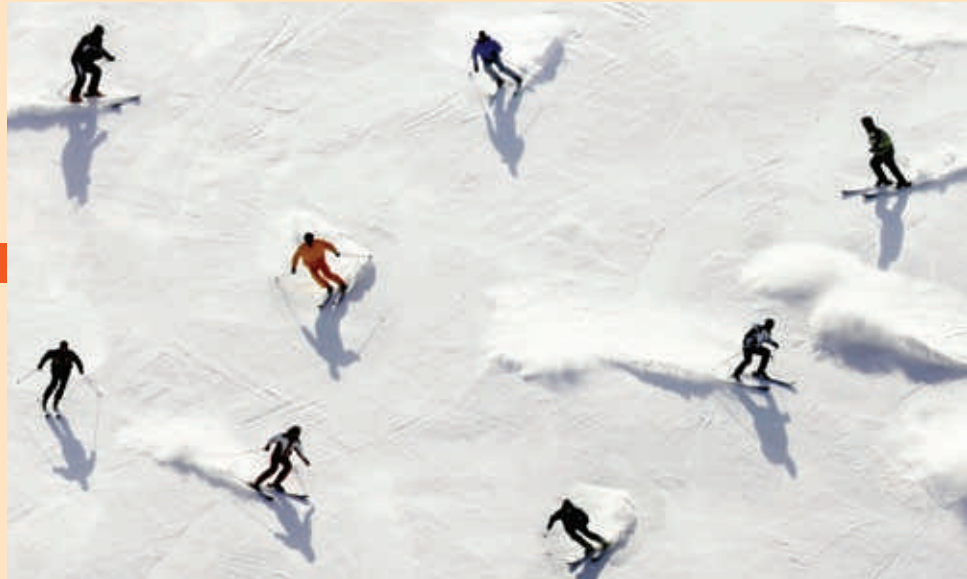


Experiments

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

Topics Appropriate for Experiments

The Classical Experiment

- Independent and Dependent Variables
- Pretesting and Posttesting
- Experimental and Control Groups
- The Double-Blind Experiment

Selecting Subjects

- Probability Sampling
- Randomization
- Matching
- Matching or Randomization?

Variations on Experimental Design

- Preexperimental Research Designs
- Validity Issues in Experimental Research

An Illustration of Experimentation

Alternative Experimental Settings

- Factorial Designs
- Web-Based Experiments
- “Natural” Experiments

Strengths and Weaknesses of the Experimental Method

Ethics and Experiments

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

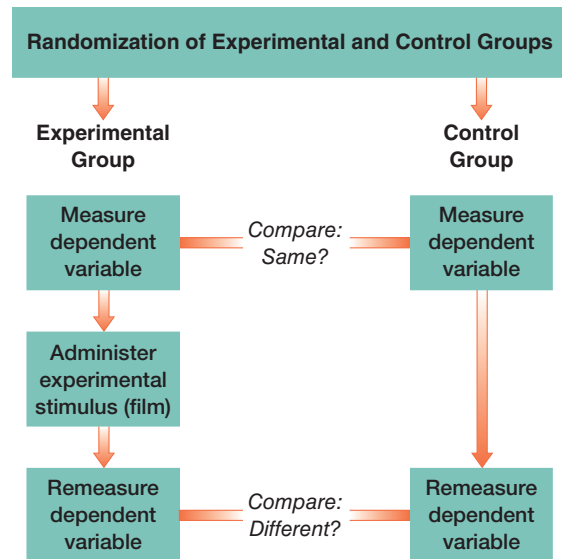


FIGURE 8-1

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 prejudice 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

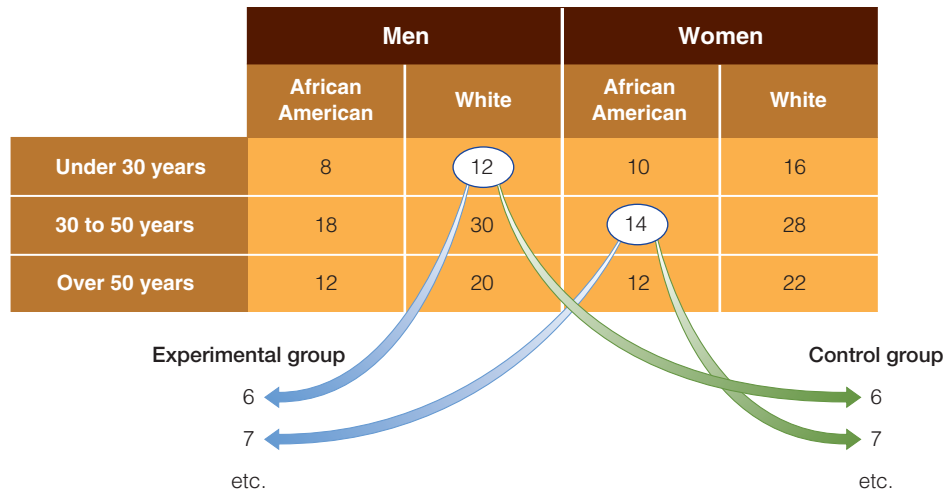
Having recruited, by whatever means, a total group of subjects, the experimenter may randomly assign those subjects to either the experimental or the control group. The researcher might accomplish such **randomization** by numbering all of the subjects serially and selecting numbers by means of a random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

**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

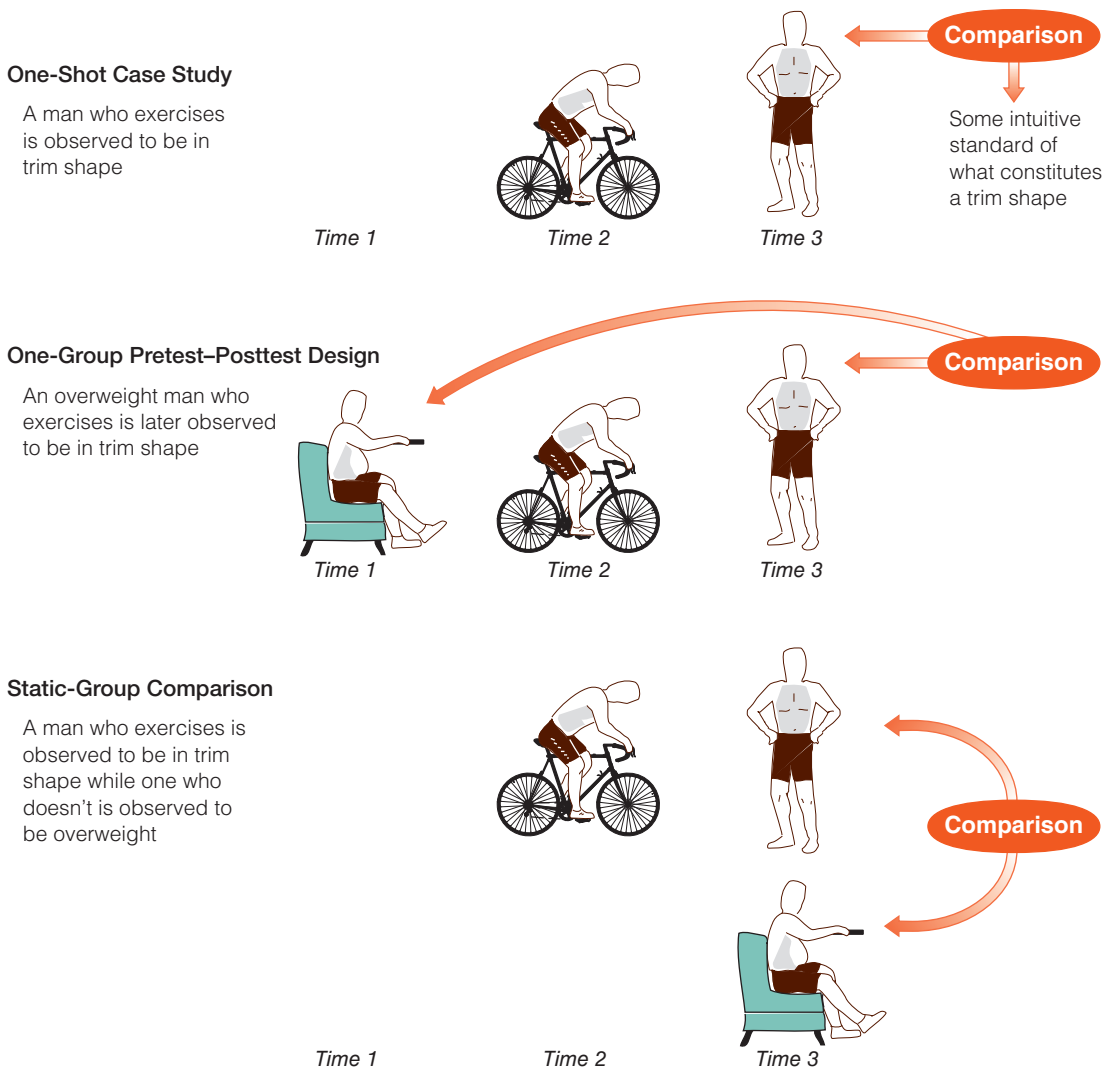


FIGURE 8-3

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.

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

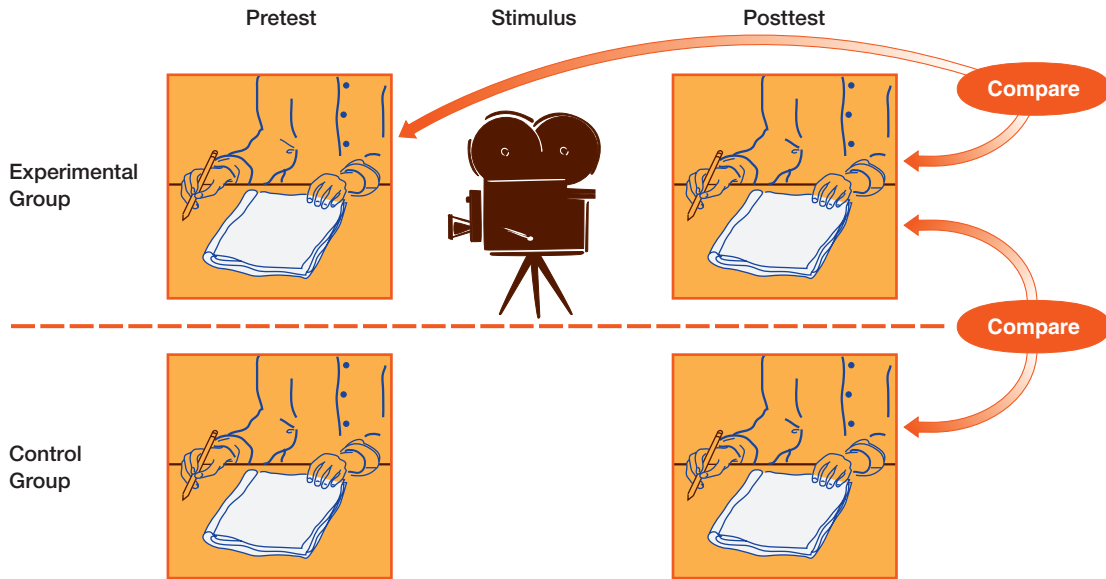


FIGURE 8-4

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?

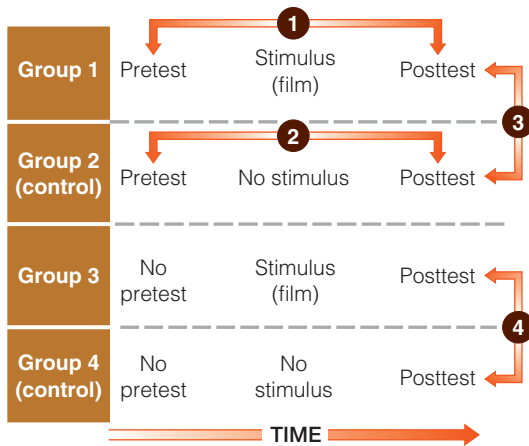
Campbell and Stanley describe four forms of this problem; I’ll present one of them as an

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 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

Topics Appropriate for Survey Research

Guidelines for Asking Questions

- Choose Appropriate Question Forms
- Make Items Clear
- Avoid Double-Barreled Questions
- Respondents Must Be Competent to Answer
- Respondents Must Be Willing to Answer
- Questions Should Be Relevant
- Short Items Are Best
- Avoid Negative Items
- Avoid Biased Items and Terms

Questionnaire Construction

- General Questionnaire Format
- Formats for Respondents

Contingency Questions

- Matrix Questions
- Ordering Items in a Questionnaire
- Questionnaire Instructions
- Pretesting the Questionnaire
- A Composite Illustration

Self-Administered Questionnaires

- Mail Distribution and Return
- Monitoring Returns
- Follow-Up Mailings
- Response Rates
- Compensation for Respondents
- A Case Study

Interview Surveys

- The Role of the Survey Interviewer
- General Guidelines for Survey Interviewing
- Coordination and Control

Telephone Surveys

- Computer-Assisted Telephone Interviewing (CATI)
- Response Rates in Interview Surveys

Online Surveys

- Online Devices
- Electronic Instrument Design
- Improving Response Rates

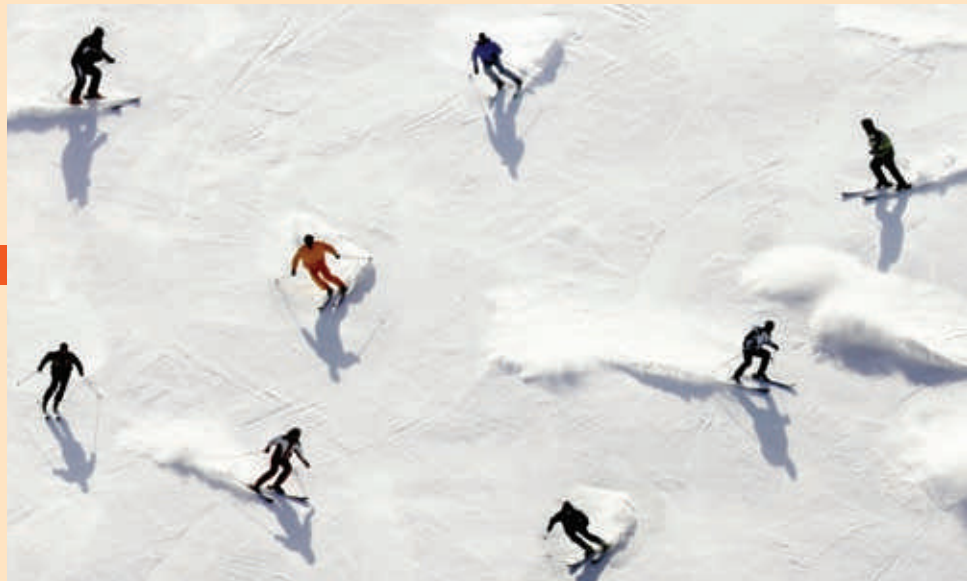
Mixed-Mode Surveys

Comparison of the Different Survey Methods

Strengths and Weaknesses of Survey Research

Secondary Analysis

Ethics and Survey Research



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 questionnaires 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 examples 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,

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

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.

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.

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:

	<i>U.S. Will Not Go to War</i>	<i>War Is Probable but Not Inevitable</i>	<i>War Is Inevitable</i>
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

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:

<i>More Support</i>	<i>Less Support</i>
“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®

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.

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®

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.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b. The police should be disarmed in America.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c. During riots, looters should be shot on sight.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
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.

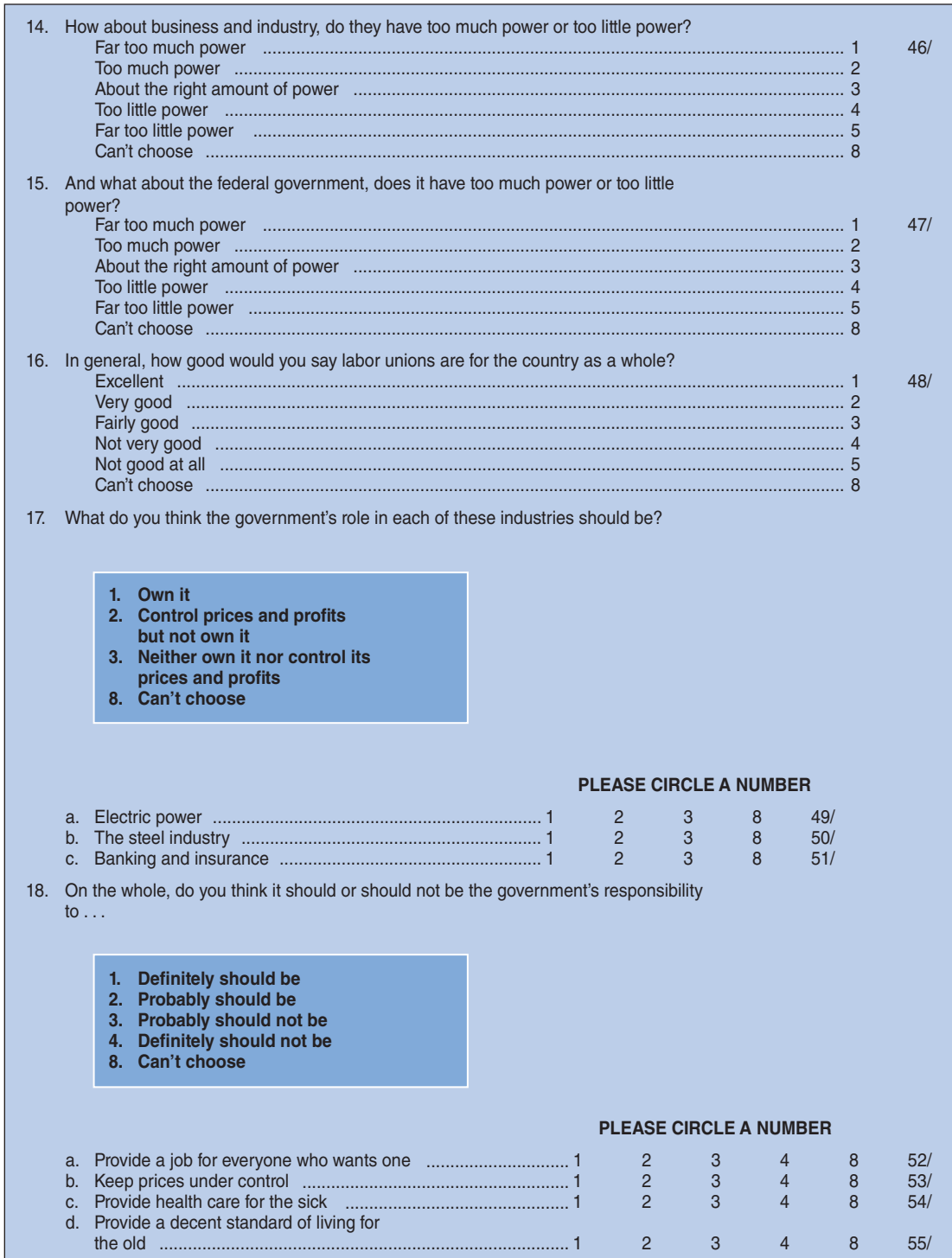


FIGURE 9-6

(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 pre-paid 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 questionnaire 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 significance 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 themselves 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 inarticulate or ambiguous to permit interpretation. However, the interviewer may be able to understand the intent of the response through the respondent'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, embarrassment, 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

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.

*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.

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.

The screenshot shows the SDA 4.0 web interface. At the top, it says "Selected Study: GSS 1972-2012 Cumulative Datafile". Below that are tabs for "Analysis", "Create Variables", and "Download Custom Subset". The "Analysis" tab is active, showing a search bar and links for "Standard Codebook" and "Codebook by Year of Interview".

The "Variable Selection" section on the left has a "Selected:" field and a "View" button. Below it are "Copy to:" buttons for "Row", "Col", "Ctrl", and "Filter". The "Mode:" section has radio buttons for "Append" and "Replace".

The main variable selection tree on the left is expanded to "Abortion". Under "Abortion", several variables are listed, including "ABANY - ABORTION IF WOMAN WANTS FOR /", "ABORCT - HEARD SUPREME COURT DECISIC", "ABPRO1 - 1ST ARGUMENT FOR ABORTION", "ABPRO2 - 2ND ARGUMENT FOR ABORTION", "ABPRO3 - 3RD ARGUMENT FOR ABORTION", "ABCON1 - 1ST ARGUMENT AGAINST ABORTI", "ABCON2 - 2ND ARGUMENT AGAINST ABORTI", and "ABCON3 - 3RD ARGUMENT AGAINST ABORTI".

The main analysis configuration panel on the right has tabs for "Tables", "Means", "Correl. matrix", "Comp. correl.", "Regression", "Logit/Probit", and "List values". The "Tables" tab is active, showing the "SDA Frequencies/Crosstabulation Program" interface. It includes fields for "Row:" (ABANY), "Column:" (RELIG), "Control:" (empty), "Selection Filter(s):" (YEAR(2012)), and "Weight:" (COMPWT - Composite weight: WTSSALL * OVERSAMP * FC). There are also "Output Options" for "Cell contents" (Percentaging: Column, Row, Total), "Simple design" (Complex, SRS), "Confidence intervals" (Level: 95 percent), "Standard error of each percent", "Design effect (defl) for each percent", "Z-statistic", "Unweighted N", and "Weighted N". Other options include "Summary statistics", "Question text", "Color coding", "Suppress table", and "Include missing data values". There is a "Title:" field and "Run the Tables" and "Clear Fields" buttons at the bottom.

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 Datasets: 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	0.0,0	1
Column	RELIG	RS RELIGIOUS PREFERENCE	1-13	0.98,99	1
Weight	COMPWT	Composite weight = WTSSALL * OVERSAMP * FORMWT	1025-11,1207		1
Filter	YEAR(2012)	GSS YEAR FOR THIS RESPONDENT	1972-2012		1

		RELIG													ROW TOTAL
Cells contain -Column percent -Weighted N		1	2	3	4	5	6	7	8	9	10	11	12	13	
		PROTESTANT	CATHOLIC	JEWISH	NONE	OTHER	BUDDHISM	HINDUISM	OTHER EASTERN	MOSLEM/ISLAM	ORTHODOX-CHRISTIAN	CHRISTIAN	NATIVE AMERICAN	INTER-NONDOMINANT	
1: YES	37.1 206.7	38.1 120.8	90.0 15.3	63.3 143.4	62.6 17.8	76.0 1.2	71.5 5.1	85.3 5.0	8.7 1.6	28.5 .8	32.2 24.5	0 0	11.5 4	42.8 836.5	
ABANY 2: NO	62.9 348.8	61.9 196.0	10.0 1.7	36.7 83.2	37.4 6.8	25.0 .4	28.5 2.0	14.7 .8	91.3 17.2	71.5 2.1	67.8 51.4	100.0 2.5	86.5 2.8	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 18.6	100.0 2.9	100.0 75.9	100.0 2.5	100.0 3.0	100.0 1,253.2	

Color coding: <0.5 <1.5 <0.5 <0.5 <1.5 <0.5 <2
in each cell: Smaller than expected Larger than expected

FIGURE 9-8

Impact of Religion on Attitude toward Abortion.

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

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
contingency question	respondent
interview	response rate
	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.