4 CHAPTER

CHAPTER OUTLINE

Choosing a Research Design

After spending long hours reading and digesting the literature in a particular research area, you have isolated a behavior that needs further investigation. You have identified some potentially important variables and probably have become familiar with the methods commonly used to measure that behavior. You may even have developed some possible explanations for the relationships that you have identified through your reading and personal experience. You are now ready to choose a research design that will allow you to evaluate the relationships that you suspect exist.

Choosing an appropriate research design is crucially important to the success of your project. The decisions you make at this stage of the research process do much to determine the quality of the conclusions you can draw from your research results. This chapter identifies the problems you must face when choosing a research design, introduces the major types of research design, and describes how each type attempts to solve (or at least cope with) these problems.

FUNCTIONS OF A RESEARCH DESIGN

Scientific studies tend to focus on one or the other of two major activities. The first activity consists of *exploratory data collection and analysis*, which is aimed at classifying behaviors within a given area of research, identifying potentially important variables, and identifying relationships between those variables and the behaviors. Such exploration is typical of the early stages of research in an area. The second activity, called *hypothesis testing*, consists of evaluating potential explanations for the observed relationships. Testable explanations allow you to predict what relationships should and should not be observed if the explanation is correct. Hypothesis testing usually begins after you have collected enough information about the behavior to begin developing supportable explanations.

CAUSAL VERSUS CORRELATIONAL RELATIONSHIPS

The relationships that you identify in these activities fall into two broad categories: causal and correlational. In a **causal relationship**, one variable directly or indirectly influences another. In other words, changes in the value of one variable directly or indirectly cause changes in the value of a second. For example, if you accidentally drop a brick on your toe, the impact of the brick will probably set off a chain of events (stimulation of pain receptors in your toe, avalanche of neural impulses traveling up your leg to the spinal cord and from there to your brain, registration of pain in your brain, involuntary scream). Although there are several intervening steps between the impact of the brick on your toe and the scream, it would be proper in this case to conclude that dropping the brick on your toe *causes* you to scream. This is because it is possible to trace an unbroken chain of physical influence running from the initial event (impact of brick on toe) to the final result (scream).

Causal relationships can be unidirectional, in which case Variable A influences Variable B but not vice versa. The impact of the brick (A) may produce a scream (B), but screaming (B) does not cause the impact of the brick on your toe (A). They also can be bidirectional, in which case each variable influences the other. Everything else being equal, reducing the amount of exercise a person gets leads to weight gain. Because of the increased effort involved, heavier people tend to exercise less. Thus, exercise influences body weight, and body weight influences exercise. Even more complex causal relationships exist, and teasing them out may require considerable ingenuity on the part of the investigator. In each case, however, one can identify a set of physical influences that ties the variables together.

Simply observing that changes in one variable tend to be associated with changes in another is not enough to establish that the relationship between them is a causal one. In a **correlational relationship**, changes in one variable accompany changes in another, but the proper tests have not been conducted to show that either variable actually influences the other. Thus, all that is known is that a relationship between them exists. When changes in one variable tend to be accompanied by specific changes in another, the two variables are said to *covary*. However, such covariation does not necessarily mean that either variable exerts an influence on the other (although it may). The number of baseball games and the number of mosquitoes tend to covary (both increase in the spring and decrease in the fall), yet you would not conclude that mosquitoes cause baseball games or vice versa.

When you first begin to develop explanations for a given behavior, knowledge of observed relationships can serve as an important guide even though you may not yet know which relationships are causal. You simply make your best guess and then develop your explanation based on the causal relationships that you think exist. The validity of your explanation will then depend in part on whether the proposed causal relationships turn out, on closer examination, to be in fact causal. Distinguishing between causal and correlational relationships is thus an important part of the research process, particularly in the hypothesis-testing phase.

Your ability to identify causal relationships and to distinguish causal from correlational relationships varies with the degree of control that you have over the variables under study. The next sections describe two broad types of research design: correlational and experimental. Both approaches allow you to identify relationships among variables, but they differ in the degree of control exerted over variables and in the ability to identify causal relationships. We begin with correlational research.

QUESTIONS TO PONDER

1. How are correlational and causal relationships similar, and how are they different?
2. Can a causal relationship be bidirectional? Explain.

CORRELATIONAL RESEARCH

In **correlational research**, your main interest is to determine whether two (or more) variables covary and, if so, to establish the directions, magnitudes, and forms of the observed relationships. The strategy involves developing measures of the variables of interest and collecting your data.

Correlational research belongs to a broader category called *nonexperimental* research, which also includes designs not specifically aimed at identifying relationships between variables. The latter type of research, for example, might seek to determine the average values and typical spread of scores on certain variables (e.g., grade point average and SAT scores) in a given population (e.g., applicants for admission to a particular university). Strictly speaking, such a study would be nonexperimental but not correlational. Our discussion here focuses on those nonexperimental methods used to identify and characterize relationships.

Correlational research involves observing the values of two or more variables and determining what relationships exist between them. In correlational research, you make no attempt to manipulate variables but observe them "as is." For example, imagine that you wished to determine the nature of the relationship, if any, between pretest anxiety and test performance in introductory psychology students on campus. On test day, you have each student rate his or her own level of pretest anxiety and, after the test results are in, you determine the test performances of those same students. Your data consist of two scores for each student: self-rated anxiety level and test score. You analyze your data to determine the relationship (if any) between these variables. Note that both anxiety level and test score are simply observed as found in each student.

In some types of correlational research, you compare the average value of some variable across preformed groups of individuals where membership in a group depends on characteristics or circumstances of the participant (such as political party affiliation, eye color, handedness, occupation, economic level, or age). For example, you might compare Democrats to Republicans on attitudes toward education. Such a study would qualify as correlational research because group membership (whether Democrat or Republican) was determined by the participants' choice of party and was not in the hands of the researcher.

Establishing that a correlational relationship exists between two variables makes it possible to predict from the value of one variable the probable value of the other variable. For example, if you know that college grade point average (GPA) is correlated with Scholastic Assessment Test (SAT) scores, then you can use a student's SAT score to predict (within limits) the student's college GPA.

When you use correlational relationships for prediction, the variable used to predict is called the *predictor variable*, and the variable whose value is being predicted is called the *criterion variable*. Whether the linkage between these variables is causal remains an open question.

An Example of Correlational Research: Cell Phone Use and Motor Vehicle Accidents

The opening vignette of Chapter 1 described the case of Bailey Goodman, the driver whose fatal crash may have resulted from distraction while texting on a cell phone. Even before texting became popular, researchers had already begun to investigate the possible dangers of cell phone use while driving. In 1997, David Redelmeier and Robert Tibshirani published a correlational study that examined the relationship between motor vehicle collisions and cell phone use. Drivers who had been involved in motor vehicle collisions that produced substantial property damage but no personal injury were recruited for the study. The cell phone records of these drivers were obtained for the day of the collision and for the preceding seven days. These records allowed Redelmeier and Tibshirani to compare the incidence of cell phone use during or just prior to the accident to its incidence at other times. They found that cell phone use "was associated with a quadrupling of the risk of a motor vehicle collision" (Redelmeier & Tibshirani, 1997, p. 455). McEvoy, Stevenson, McCartt, and colleagues (2005) obtained nearly identical results in a similar study involving drivers whose accidents had resulted in hospital attendance.

Assessing the Redelmeier and Tibshirani Study What qualifies Redelmeier and Tibshirani's study as a correlational study? In their study, cell phone usage at the time of the accident and at other times was simply recorded as found. No attempt was made to manipulate variables in order to observe any potential effects of those variables.

Behavior Causation and the Correlational Approach

Given the results obtained by Redelmeier and Tibshirani's (1997) study and by McEvoy et al. (2005), you might be tempted to conclude that using a cell phone while driving *causes* motor vehicle accidents. However, this conclusion that a causal relationship exists is inappropriate even though the relationship appears compelling. Two obstacles stand in the way of drawing clear causal inferences from correlational data: the third-variable problem and the directionality problem.

The Third-Variable Problem To establish a causal relationship between two variables, you must be able to demonstrate that variation in one of the observed variables could only be due to the influence of the other observed variable. In the example, you want to show that variation in the cell phone use while driving causes variation in the risk of a motor vehicle accident. However, because the drivers (and not the researchers) chose whether or not to use a cell phone while driving, it is possible that the observed relationship between cell phone use and the risk of a motor vehicle accident may actually be due to the influence of a third variable. For example, drivers may be more likely to talk on a cell phone while driving when they are distressed about some personal matter. This distress might also compromise a driver's ability to focus on his or her driving, thus leading to an increased risk of an accident. Although far-fetched, such a possibility cannot be ruled out in the studies cited.

The possibility that correlational relationships may result from the action of an unobserved "third variable" is called the **third-variable problem**. This unobserved variable may influence both of the observed variables (e.g., cell phone use and the likelihood of having a motor vehicle accident), causing them to vary together even though no direct relationship exists between them. The two observed variables thus may be strongly correlated even though neither variable causes changes in the other.

To resolve the third-variable problem, you must examine the effects of each potential third variable to determine whether it does, in fact, account for the observed relationship. Techniques to evaluate and statistically control the effects of such variables are available (see Chapter 15).

The Directionality Problem A second reason why it is hazardous to draw causal inferences from correlational data is that, even when a direct causal relationship exists, the direction of causality is sometimes difficult to determine. This difficulty is known as the **directionality problem**.

The directionality problem does not apply to the cell phone studies as it is not possible that having a motor vehicle accident could cause a person to be using a cell phone in the minutes or seconds preceding the accident. However, it can pose a problem for some studies. For example, Anderson and Dill (2000) found a positive relationship between level of aggression (as self-reported by students in their questionnaires) and the amount of exposure to violent video games. You might be tempted to conclude that students become more aggressive from playing violent video games, but it seems just as reasonable to turn the causal arrow around. Perhaps finding gratification in aggressive behavior leads to a preference for playing violent video games.

Why Use Correlational Research?

Given the problems of interpreting the results of correlational research, you may wonder why you would want to use this approach. However, correlational research has a variety of applications, and there are many reasons to consider using it. In this section, we discuss three situations in which a correlational approach makes good sense.

Gathering Data in the Early Stages of Research During the initial, exploratory stage of a research project, the correlational approach's ability to identify potential causal relationships can provide a rich source of hypotheses that later may be tested experimentally. Consider the following example.

Niko Tinbergen (1951) became interested in the behavior of the three-spined stickleback, a fish that inhabits the bottoms of sandy streams in Europe. Observing sticklebacks in their natural habitat, Tinbergen found that, during the spring, the male stickleback claims a small area of a streambed and builds a cylindrically shaped nest at its center. At the same time, the male's underbelly changes from the usual dull color to a bright red, and the male begins to drive other males from the territory surrounding the nest. Female sticklebacks lack this coloration and are not driven away by the males.

These initial observations were purely correlational and as such do not allow one to draw firm conclusions with respect to cause and effect. The observations showed that the defending male's behavior toward an intruding stickleback is correlated with the intruder's physical characteristics, but which characteristics actually determine whether or not an attack will occur? Certainly many cues, such as the male's red coloration, his shape, or even perhaps his odor, could be responsible. However, these cues always appeared and disappeared together (along with the fish to which they belonged). So there was no way, through correlational study alone, to determine whether the red coloration was the actual cause of the defensive behavior or merely an ineffective correlate.

To disentangle these variables, Tinbergen (1951) turned to the experimental approach. He set up an artificial stream in his laboratory and brought in several male sticklebacks. The fish soon adapted to the new surroundings, setting up territories and building nests. Tinbergen then constructed a number of models designed to mimic several characteristics of male sticklebacks. These models ranged from one that faithfully duplicated the appearance (but not the smell) of a real stickleback to one that was just a gray disk (Figure 4-1). Some of the models included red coloration, and some did not.

When the realistic model with a red underbelly was waved past a male stickleback in the artificial stream, the male immediately tried to drive it away. Odor obviously was not necessary to elicit defensive behavior. However, Tinbergen (1951) soon discovered that almost any model with red color elicited the response. The only requirements were that the model include an eyespot near the top and that the red color appear below the eyespot.

FIGURE 4-1 Stimuli used by Tinbergen to follow up on initial observations made in the field: N, neutral underbelly; R, red underbelly.
SOURCE: Tinbergen, 1951; reprinted with permission.

By manipulating factors such as color and shape, Tinbergen (1951) could experimentally identify the factors that were necessary to elicit the behavior. The earlier, correlational research conducted in a naturalistic (and therefore poorly controlled) setting had paved the way for the more definitive research that followed.

Inability to Manipulate Variables In an experimental design, variables are manipulated to determine their effects on other variables. A second reason for choosing a correlational design over an experimental one is that manipulating the variables of interest may be impossible or unethical (see Chapter 7 for a discussion of ethics). For example, imagine that you were interested in determining whether psychopathic personality develops when a child is raised by cold, uncaring parents. To establish a clear causal connection between the parents' behavior toward the child and psychopathic personality, you would have to conduct an experiment in which the parents' behavior was manipulated by assigning infants at random to be raised by either normal parents or cold, uncaring ones. However, this experiment would be impossible to carry out (who would allow their child to participate in such an experiment?) and, because of its potential for inflicting serious harm on the child, unethical as well. In such cases, a correlational design may be the only practical and ethical option.

Relating Naturally Occurring Variables A third situation in which you may choose a correlational research design over an experimental design is one in which you want to see how naturally occurring variables relate in the real world. Such information can be used to make useful predictions even if the reasons for the discovered relationships are not clear. High school GPA, scores on the SAT, class rank, and scores on the Nelson–Denny reading comprehension test correlate well with each other and with performance in college. Knowledge of these relationships has been used to predict college success. Certain theoretical views also may lead to predictions about which real-world variables should be correlated with which. These predictions can be tested by using a correlational design.

QUESTIONS TO PONDER

1. What are the defining features of correlational research?
2. Why is it inappropriate to draw causal inferences from correlational data?
3. Under what conditions is correlational research preferred over experimental research?

EXPERIMENTAL RESEARCH

Unlike correlational research, experimental research incorporates a high degree of control over the variables of your study. This control, if used properly, permits you to establish causal relationships among your variables. This section describes the defining characteristics of experimental research and explains how these characteristics enable us to identify causal relationships in data.

Characteristics of Experimental Research

Experimental research has two defining characteristics: manipulation of one or more independent variables and control over extraneous variables. Be sure that you understand these concepts, described as follows, because they are central to understanding experimental research.

Manipulation of Independent Variables An **independent variable** is a variable whose values are chosen and set by the experimenter. (Another way to look at it is that the value of the independent variable is independent of the participant's behavior.) We call these set values the *levels* of the independent variable. For example, imagine that you want to determine how sleep deprivation affects a person's ability to recall previously memorized material. To examine this relationship, you might assign participants to one of three groups defined by the number of hours of sleep deprivation: 0 hours (rested), 24 hours, and 48 hours. These three amounts would constitute the three levels of sleep deprivation, your independent variable.

To *manipulate* your independent variable, you must expose your participants to at least two levels of that variable. The specific conditions associated with each level are called the **treatments** of the experiment. Depending on the design of your experiment, the independent variable may be manipulated by exposing a different group of participants to each treatment or by exposing each participant to all the treatments in sequence. By manipulating the independent variable, you hope to show that changes in the level of the independent variable *cause* changes in the behavior being recorded.

The variable whose value you observe and measure in experimental designs is called the **dependent variable** (or *dependent measure*). If a causal relationship exists, then the value of the dependent variable depends, at least to some extent, on the level of the independent variable. (Its value also depends on other factors such as participant characteristics.) Another way to think about the dependent variable is that its value depends on the behavior of the participant, rather than being set by the experimenter.

Manipulating an independent variable can be as simple as exposing one group of participants to some treatment (e.g., distracting noises) and another group of participants to the absence of the treatment (no distracting noise). In this most basic of experimental designs, the group receiving the treatment is called the **experimental group** and the other group the **control group**. The control group is treated exactly like the experimental group except that it is not exposed to the experimental treatment. The performance of the participants in the control group provides a baseline of behavior against which the behavior of the participants in the experimental groups is compared.

Although all experiments present at least two levels of the independent variable, many do not include a no-treatment control group. A clinical study, for example, might compare a standard therapy with a new, experimental therapy of unknown effectiveness. Administering the standard therapy to the control group ensures that even the participants who do not receive the experimental treatment do not go untreated for their disorder. In both cases, the behavior of participants in the control group provides a baseline against which to compare the behavior of participants in the experimental group.

More complex experiments can be conducted using more levels of the independent variable, several independent variables, and several dependent variables. You also can choose to expose a single group, or even a single participant, to several levels of an independent variable.

Control Over Extraneous Variables The second characteristic of experimental research is control over extraneous variables. **Extraneous variables** are those that may affect the behavior that you wish to investigate but are not of interest for the present experiment. For example, you may be interested in determining how well a new anxiety therapy (experimental group), compared with an existing therapy (control group), affects test anxiety in anxious students. If some of your participants show up for the experiment drunk, their degree of intoxication becomes an extraneous variable. This would be especially problematic if more drunk students ended up in one group than in the other.

If allowed to vary on their own, extraneous variables can produce uncontrolled changes in the value of the dependent variable, with two rather nasty possible consequences. First, uncontrolled variability may make it difficult or impossible to detect any effects of the independent variable. (In our example, the effects of the therapy could be buried under the effects of the alcohol.) Second, uncontrolled variability may produce chance differences in behavior across the levels of the independent variable. These differences could make it appear as though the independent variable produced effects when it did not (the therapy would appear to work even though the real effect came from the alcohol). To identify clear causal relationships between your independent and dependent variables, you must control the effects of extraneous variables.

You have two ways to control these effects. The first way is simply to *hold extraneous variables constant*. If these variables do not vary over the course of your experiment, they cannot cause uncontrolled variation in your dependent variable. In the test anxiety experiment, for example, you might want to make sure that all your participants are sober (or at least intoxicated to the same degree). In fact, to the degree possible, you would want to make sure that all treatments are *exactly* alike, except for the level of the independent variable.

The second way to deal with extraneous variables is to *randomize their effects across treatments*. This technique deals with the effects of extraneous variables that cannot be held constant or, for reasons that will be explained later, should not be held constant. In an experiment assessing the effect of sleep deprivation on memory, for example, it may not be possible to ensure that all your participants have had identical amounts of sleep deprivation (some may have slept better than others the day before your experiment began) or that their recall abilities are equivalent. The idea is to distribute the effects of these differences across treatments in such a way that they tend to even out and thus cannot be mistaken for effects of the independent variable.

For statistical reasons, one of the better ways to accomplish this goal is to use random assignment of subjects to treatments. With **random assignment**, you assign participants to treatments randomly by picking their names out of a hat, for example. (In practice, one does not use names in a hat.) A table of random numbers can be used to assign subjects to treatment conditions randomly. Random assignment does not *guarantee* that the effects of extraneous variables will be distributed evenly across treatments, but it usually works reasonably well; better yet, it allows you to use inferential statistics to evaluate the probability with which chance alone could have produced the observed differences. (We discuss the logic underlying inferential statistics in Chapter 14.) Other techniques to deal with uncontrolled extraneous variables are also available. We describe these in later chapters that cover specific design options.

However it is done, control over extraneous variables is crucial to establishing clear causal relationships between your variables. By controlling variables that might affect your dependent variable, you rule them out as possible alternative explanations for your results.

An Example of Experimental Research: Cell Phone Use While Driving

As an illustration of experimental research, consider a follow-up study conducted by David Strayer and Frank Drews (2007), whose earlier research we summarized briefly in Chapter 1. The earlier research had shown that cell phone use seriously impairs performance in a simulated driving task. In the 2007 study, Strayer and Drews tested the hypothesis that "cell-phone conversations impair driving by inducing a form of inattention blindness in which drivers fail to see objects in their driving environment when they are talking on a cell phone" (Strayer & Drews, 2007, p. 128). Participants drove in a simulator that closely resembled the interior of a Ford Crown Victoria and offered a realistic view of a simulated road through the front and side windows. A video system monitored the driver's eye movements. In one experiment, some participants drove while conversing on a hands-free cell phone; others drove without conversing. (Participants were randomly assigned to the conditions.) After completing the driving course, the drivers were tested for recognition of objects in the scenery they had "passed" along the way. The analysis focused on those objects on which the drivers' eyes had fixated during the drive. Those drivers who had been conversing on the cell phone while driving recognized significantly fewer objects than those who had been driving without conversing. Based on this finding and others from the study, Strayer and Drews concluded that "these data support an inattention-blindness interpretation wherein the disruptive effects of cell-phone conversations on driving are due in large part to the diversion of attention from driving to the phone conversation." (p. 128).

Assessing the Strayer and Drews Experiment Have you identified the features of the Strayer and Drews (2007) experiment that qualify it as a true experiment? If you have not done so yet, do it now before you read the next paragraphs.

A crucial element of every true experiment is the manipulation of at least one independent variable. What is the independent variable in the Strayer and Drews (2007) study? If you said that the presence or absence of a cell phone conversation while driving was the independent variable, you are correct. Note that the value of the independent variable to which a given participant was exposed (cell phone conversation or no conversation) was assigned by the experimenters; it was not chosen by the participant.

A second crucial element in an experiment is measuring a dependent variable. Can you identify the dependent variable in Strayer and Drews' (2007) experiment? If you said that the ability to recall details about the objects on which the driver fixated was the dependent variable, you are correct. Notice that Strayer and Drews were looking for changes in the value of the dependent variable relating to changes in the value of the independent variable.

A third crucial element of an experiment is control over extraneous variables. Were extraneous variables controlled in the Strayer and Drews (2007) experiment and, if so, how? The answer to the first part of this question is yes, and if you examine the design of the study carefully, you will see that extraneous variables were controlled using both methods described earlier. First, several extraneous variables were held constant across treatments. For example, all drivers used the same simulator and saw identical scenery along the "route." And other than the use of a cell phone or not, both groups of participants received the same treatment. Second, the participants were assigned to their treatments randomly, not according to some behavior or characteristic of the participants. This design ensured that any remaining uncontrolled differences in the participants would tend to be distributed evenly between the two treatments. As a result, the investigators could be reasonably sure that any differences found between treatments in the values of the dependent measures were caused by the difference in treatments—that is, by the difference between holding a conversation on a cell phone while driving and not doing so.

Strengths and Limitations of the Experimental Approach

The great strength of the experimental approach is its ability to identify and describe causal relationships. This ability is not shared by the correlational approach. Whereas the correlational approach can tell you only that changes in the value of one variable tend to accompany changes in the value of a second variable, the experimental approach can tell you whether changes in one variable (the independent variable) actually caused changes in the other (the dependent variable).

Despite its power to identify causal relationships, the experimental approach has limitations that restrict its use under certain conditions. The most serious limitation is that you cannot use the experimental method if you cannot manipulate your hypothesized causal variables. For example, studies of personality disorders must use correlational approaches to identify possible causal relationships. Exposing people to various nasty conditions in order to identify which of those conditions cause personality disorders is not ethical.

A second limitation of the experimental approach entails the tight control over extraneous factors required to clearly reveal the effects of the independent variable. Such control tends to reduce your ability to apply your findings to situations that differ from the conditions of your original experiment. A rather unpleasant trade-off exists in experimental research: As you increase the degree of control that you exert over extraneous variables (and thus your ability to establish causal relationships), you decrease your ability to assess the generality of any relationships you uncover. For example, in the Strayer and Drews (2005) experiment, extraneous variables such as simulated traffic and scenery were controlled. However, this control may limit the generality of their results because it is possible that different results would be obtained using other traffic scenarios that are, for example, more or less demanding. (We discuss the problem of generality more fully later in the chapter.)

Experiments Versus Demonstrations

One kind of research design resembles an experiment but lacks one of the crucial features of a true experiment, an independent variable. This design, called a **demonstration**, exposes a group of subjects to one (and only one) treatment condition. Remember, a true experiment requires exposing subjects to at least two treatments. Whereas a true experiment shows the effect of manipulating an independent variable, a demonstration simply shows what happens under a specified set of conditions. To conduct a demonstration, you simply expose a single group to a particular treatment and measure the resulting behavior.

Demonstrations can be useful because they show that, under such-and-such conditions, *this* happens and not *that*. However, demonstrations are not experiments and thus do not show causal relationships. This fact is sometimes overlooked as the following example shows.

In his book *Subliminal Seduction* (1973), Wilson Bryan Key reported a study in which the participants looked at a Gilbey's Gin advertisement that allegedly had subliminal sexual messages embedded within it. The most prominent subliminal message was the word "SEX" spelled out in the bottom three ice cubes in the glass to the right of a bottle of gin (Key, 1973).

Key (1973) reported that the ad was tested "with over a thousand subjects" (the details of the study were not given). According to Key, 62% of the male and female participants reported feelings of sexual arousal in response to the ad. Key concluded that the subliminal messages led to sexual arousal. Key asserted that advertisers capitalize on these subliminal messages to get you to buy their products.

Are you convinced of the power of subliminal messages by this demonstration? If you said you are not convinced, good for you! The fact that 62% of the participants reported arousal is *not* evidence that the subliminal messages caused the arousal, no matter how many participated. All you know from this demonstration is that under the conditions tested, the advertisement evoked reports of arousal in a fair proportion of the participants. You do not learn the cause.

In fact, several plausible alternatives can be offered to the explanation that the arousal was caused by subliminal perception. For example, an advertisement for alcohol may lead participants to recall how they feel when under the influence or may conjure up images of having fun at a party. As the demonstration was reported, you cannot tell which of the potential explanations is valid. What would you have to do to fully test whether subliminal messages (such as the ones in the Gilbey's Gin ad) actually lead to sexual arousal? Give this question some thought before continuing.

To test whether subliminal messages caused the arousal, you need to add a control group and randomly assign participants to groups. Participants in this control group would see the same Gilbey's Gin ad but without the subliminal messages. If 62% of the participants in the "subliminal" group were aroused but only 10% in the control group were aroused, then you could reasonably conclude that the subliminal messages caused the arousal. A different conclusion would be drawn if 62% of the participants in *both* groups reported arousal. In this case, you would have to conclude that the subliminal messages were ineffective. The fact that the ad leads to reports of sexual arousal (as shown by the demonstration) would have to be explained by some other factor. By the way, most of the controlled, scientific research on subliminal perception shows little or no effect of subliminal messages on behavior.

QUESTIONS TO PONDER

1. What are the characteristics of experimental research?
2. What is the relationship between an independent and a dependent variable in an experiment?
3. How do extraneous variables affect your research?
4. What can be done to control extraneous variables?
5. How does a demonstration differ from a true experiment?
6. What is the value of doing a demonstration?

INTERNAL AND EXTERNAL VALIDITY

Whether the general design of your study is experimental or correlational, you need to consider carefully two important but often conflicting attributes of any design: internal and external validity. In this section, we define these concepts and briefly discuss the factors that you should consider relating to internal and external validity when choosing a research design.

Internal Validity

Much of your research will be aimed at testing the hypotheses you developed long before you collected any data. The ability of your research design to adequately test your hypotheses is known as its **internal validity** (Campbell & Stanley, 1963). Essentially, internal validity is the ability of your design to test the hypothesis that it was designed to test.

In an experiment, this means showing that variation in the independent variable, and only the independent variable, caused the observed variation in the dependent variable. In a correlational study, it means showing that changes in the value of your criterion variable relate solely to changes in the value of your predictor variable and not to changes in other, extraneous variables that may have varied along with your predictor variable.

Internal validity is threatened to the extent that extraneous variables can provide alternative explanations for the findings of a study, or as Huck and Sandler (1979) call them, *rival hypotheses*. As an example, imagine that an instructor wants to know whether a new teaching method works better than the traditional method used with students in an introductory psychology course. The instructor decides to answer this question by using the new method to teach her morning section of introductory psychology and using the traditional method to teach her afternoon section. Both sections will use the same text, cover the same material, and receive the same tests. The effectiveness of the two methods will be assessed by comparing the average scores achieved on the test by the two sections. Now, imagine that the instructor conducts the study and finds that the section receiving the new method receives a substantially higher average grade than the section receiving the traditional method. She concludes that the new method is definitely better for teaching introductory psychology. Is she justified in drawing this conclusion?

The answer, as you probably suspected, is no. Several rival hypotheses cannot be eliminated by the study, explanations at least as credible as the instructor's view that the new method was responsible for the observed improvement in average grade. Consider the following rival hypotheses:

1. The morning students did better because they were "fresher" than the afternoon students.
2. The morning students did better because their instructor was "fresher" in the morning than in the afternoon.
3. The instructor expected the new method to work better and thus was more enthusiastic when using the new method than when using the old one.
4. Students who registered for the morning class were more motivated to do well in the course than those who registered for the afternoon class.

These rival hypotheses do not exhaust the possibilities; perhaps you can think of others. Because the study was not designed to rule out these alternatives, there is no way to know whether the observed difference between the two sections in student performance was due to the difference in teaching methods, instructor enthusiasm, alertness of the students, or other factors whose levels differed across the sections. Whenever two or more variables combine in such a way that their effects cannot be separated, a **confounding** of those variables has occurred. In the teaching study, teaching method is confounded by all those variables just listed and more. Such a study lacks internal validity.

Confounding, although always a matter of concern, does not necessarily present a serious threat to internal validity. Confounding is less problematic when the confounding variable is known to have little or no effect on the dependent or criterion variable or when its known effect can be taken into account in the analysis. For example, in the teaching study, it may be possible to eliminate concern about the difference in class meeting times by comparing classes that meet at different times but use the same teaching method. Such data may show that meeting time has only a small effect that can be ignored. If meeting time had a larger effect, you could arrange your study of teaching method so that the effect of meeting time would tend to make the new teaching method appear worse than the standard one, thus biasing the results against your hypothesis. If your results still favored the new teaching method, that outcome would have occurred despite the confounding rather than because of it. Thus, a study may include confounding and still maintain a fair degree of internal validity if the effects of the confounding variable in the situation under scrutiny are known.

This is fortunate because it is often impossible to eliminate all sources of confounding in a study. For example, the instructor in our example might have attempted to eliminate confounding by having students randomly assigned to two sections meeting simultaneously. This would certainly eliminate those sources of confounding related to any difference in the time at which the sections met, but now it would be impossible for the instructor to teach both classes. If a second instructor is recruited to teach one of the sections using the standard method, this introduces a new source of confounding in that the two instructors may not be equivalent in a number of ways that could affect class performance. Often the best that can be done is to substitute what you believe to be less serious threats to internal validity for the more serious ones.

Threats to Internal Validity Confounding variables occur in both experimental and correlational designs, but they are far more likely to be a problem in the latter, in which tight control over extraneous variables is usually lacking. Campbell and Stanley (1963) identify seven general sources of confounding that may affect internal validity: history, maturation, testing, instrumentation, statistical regression, biased selection of subjects, and experimental mortality (Table 4-1).

History may confound studies in which multiple observations are taken over time. Specific events may occur between observations that affect the results. For example, a study of the effectiveness of an advertising campaign against drunk driving might measure the number of arrests for drunk driving immediately before and after the campaign. If the police institute a crackdown on drunk driving at the same time that the advertisements air, this event will destroy the internal validity of your study.

Maturation refers to the effect of age or fatigue. Performance changes observed over time due to these factors may confound those due to the variables being studied. You might, for example, assess performance on a proofreading task before and after

TABLE 4-1 Factors Affecting Internal Validity

FACTOR	DESCRIPTION
History	Specific events other than the treatment occur between observations
Maturation	Performance changes due to age or fatigue confound the effect of treatment
Testing	Testing prior to the treatment changes how subjects respond in posttreatment testing
Instrumentation	Unobserved changes in observer criteria or instrument calibration confound the effect of the treatment
Statistical regression	Subjects selected for treatment on the basis of their extreme scores tend to move closer to the mean on retesting
Biased selection of subjects	Groups of subjects exposed to different treatments are not equivalent prior to treatment
Experimental mortality	Differential loss of subjects from the groups of a study results in nonequivalent groups

some experimental manipulation. Decreased performance on the second proofreading assessment may be due to fatigue rather than to any effect of your manipulation.

Testing effects occur when a pretest sensitizes participants to what you are investigating in your study. As a consequence, they may respond differently on a posttreatment measure than if no pretest were given. For example, if you measure participants' racial attitudes and then manipulate race in an experiment on person perception, participants may respond to the treatment differently than if no such pretest of racial attitudes was given.

In *instrumentation*, confounding may be introduced by unobserved changes in criteria used by observers or in instrument calibration. If observers change what counts as "verbal aggression" when scoring behavior under two experimental conditions, any apparent difference between those conditions in verbal aggression could be due as much to the changed criterion as to any effect of the independent variable. Similarly, if an instrument used to record activity of rats in a cage becomes more (or less) sensitive over time, it becomes impossible to tell whether activity is really changing or just the ability of the instrument to detect activity.

Statistical regression threatens internal validity when participants have been selected based on extreme scores on some measure. When measured again, scores will tend to be closer to the average in the population. Thus, if students are targeted for a special reading program based on their unusually low reading test scores, they will tend to do better, on average, on retesting even if the reading program has no effect.

Biased selection of subjects threatens internal validity because subjects may differ initially in ways that affect their scores on the dependent measure. Any influence of the independent variable on scores cannot be separated from the effect of the preexisting bias. This problem typically arises when researchers use preexisting groups in their studies rather than assigning subjects to groups at random. For example, the effect of a program designed to improve worker job satisfaction might be evaluated by administering the program to workers at one factory (experimental group) and then comparing the level of job satisfaction of those workers to that of workers at another factory where the program was not given (control group). If workers given the job satisfaction program indicate more satisfaction with their jobs, is it due to the program or to preexisting differences between the two groups? There is no way to tell.

Finally, *experimental mortality* refers to the differential loss of participants from groups in a study. For example, imagine that some people drop out of a study because of frustration with the task. A group exposed to difficult conditions is more likely to lose its frustration-intolerant participants than one exposed to less difficult conditions. Any differences between the groups in performance may be due as much to the resulting difference in participants as to any difference in conditions.

Enhancing Internal Validity The time to be concerned with internal validity is during the design phase of your study. During this phase, you should carefully plan which variables will be manipulated or observed and recorded, identify any plausible rival hypotheses not eliminated in your initial design, and redesign so as to eliminate those that seriously threaten internal validity. Discovering problems with internal validity after you have run your study is too late. A poorly designed study cannot be fixed later on.

External Validity

A study has **external validity** to the degree that its results can be extended (generalized) beyond the limited research setting and sample in which they were obtained. A common complaint about research using white rats or college students and conducted under the artificial conditions of the laboratory is that it may tell us little about how white rats and college sophomores (let alone animals or people in general) behave under the conditions imposed on them in the much richer arena of the real world.

The idea seems to be that all studies *should* be conducted in such a way that the findings can be generalized immediately to real-world situations and to larger populations. However, as Mook (1983) notes, it is a fallacy to assume "that the purpose of collecting data in the laboratory is to predict real-life behavior in the real world" (p. 381). Mook points out that much of the research conducted in the laboratory is designed to determine one of the following:

1. Whether something *can* happen, rather than whether it typically *does* happen
2. Whether something we specify *ought* to happen (according to some hypothesis) under specific conditions in the lab *does* happen there under those conditions
3. What happens under conditions not encountered in the real world

In each of these cases, the objective is to gain insight into the underlying mechanisms of behavior rather than to discover relationships that apply under normal conditions in the real world. It is this understanding that generalizes to everyday life, not the specific findings themselves.

Threats to External Validity In Chapter 1, we distinguished between basic research, which is aimed at developing a better understanding of the underlying mechanisms of behavior, and applied research, which is aimed at developing information that can be directly applied to solve real-world problems. The question of external validity may be less relevant in basic research settings that seek theoretical reasons to determine what will happen under conditions not usually found in natural settings or that examine fundamental processes expected to operate under a wide variety of conditions. The degree of external validity of a study becomes more relevant when the findings are expected to be applied directly to real-world settings. In such studies, external validity is affected by several factors. Using highly controlled laboratory settings (as opposed to naturalistic settings) is one such factor. Data obtained from a tightly controlled laboratory may not generalize to more naturalistic situations in which behavior occurs. Other factors that affect external validity, as discussed by Campbell and Stanley (1963), are listed and briefly described in Table 4-2. Many of these threats to external validity are discussed in later chapters, along with the appropriate research design.

TABLE 4-2 Factors Affecting External Validity

FACTOR	DESCRIPTION
Reactive testing	Occurs when a pretest affects participants' reaction to an experimental variable, making those participants' responses unrepresentative of the general population
Interactions between participant selection biases and the independent variable	Effects observed may apply only to the participants included in the study, especially if they are unique to a group (such as college sophomores rather than a cross section of adults)
Reactive effects of experimental arrangements	Refers to the effects of highly artificial experimental situations used in some research and the participant's knowledge that he or she is a research participant
Multiple treatment interference	Occurs when participants are exposed to multiple experimental treatments in which exposure to early treatments affects responses to later treatments

Internal Versus External Validity

Although you should strive to achieve a high degree of both internal and external validity in your research, in practice you will find that the steps you take to increase one type of validity tend to decrease the other. For example, a tightly controlled laboratory experiment affords you a relatively high degree of internal validity. Your findings, however, may not generalize to other samples and situations; thus, external validity may be reduced. Often the best that you can do is reach a compromise on the relative amounts of internal and external validity in your research.

Whether internal or external validity is more important depends on your reasons for conducting the research. If you are most interested in testing a theoretical position (as is often the case in basic research), you might be more concerned with internal than external validity and hence conduct a tightly controlled laboratory experiment. However, if you are more concerned with applying your results to a real-world problem (as in applied research), you might take steps to increase the external validity while attempting to maintain a reasonable degree of internal validity. These issues need to be considered at the time when you design your study.

As just mentioned, the setting in which you conduct your research strongly influences the internal and external validity of your results. The kinds of setting available and the issues that you should consider when choosing a research setting are the topics that we take up next.

QUESTIONS TO PONDER

1. What is internal validity, and why is it important?
2. What factors threaten internal validity?
3. How do confounding variables threaten internal validity, and how can they be avoided?
4. What is external validity, and when is it important to have high levels of external validity?
5. How do internal and external validity relate to one another?

RESEARCH SETTINGS

In addition to deciding on the design of your research, you also must decide on the setting in which you conduct your research. Your choice of setting is affected by the potential costs of the setting, its convenience, ethical considerations, and the research question that you are addressing.

The two research settings open for psychological research are the laboratory and the field. For this discussion, the term *laboratory* is used in a broad sense. A laboratory is any research setting that is artificial relative to the setting in which the behavior naturally occurs. This definition is not limited to a special room with special equipment for research. A laboratory can be a formal lab, but it also can be a classroom, a room in the library, or a room in the student union building. In contrast, the *field* is the setting in which the behavior under study naturally occurs.

Your decision concerning the setting for your research is an important one, so you must be familiar with the relative advantages and disadvantages of each.

The Laboratory Setting

If you choose to conduct your research in a *laboratory setting*, you gain important control over the variables that could affect your results. The degree of control depends on the nature of the laboratory setting. For example, if you are interested in animal learning, you can structure the setting to eliminate virtually all extraneous variables that could affect the course of learning. This is what Ivan Pavlov did in his investigations of classical conditioning. Pavlov exposed dogs to his experimental conditions while the dogs stood in a sound-shielded room. The shielded room permitted Pavlov to investigate the impact of the experimental stimuli free from any interfering sounds. Like Pavlov, you can control important variables within the laboratory that could affect the outcome of your research.

Complete control over extraneous variables may not be possible in all laboratory settings. For example, if you were administering your study to a large group of students in a psychology class, you could not control all the variables as well as you might wish (students may arrive late, or disruptions may occur in the hallway). For the most part, the laboratory affords more control over the research situation than does the field.

Simulation: Re-creating the World in the Laboratory When you choose the laboratory as your research setting, you gain control over extraneous variables that could affect the value of your dependent variable. However, you make a trade-off when choosing the laboratory. Although you gain better control over variables, your results may lose some generality (the ability to apply your results beyond your specific laboratory conditions). If you are concerned with the ability to generalize your results, as well as with controlling extraneous variables, consider using a **simulation**. In a simulation, you attempt to re-create (as closely as possible) a real-world situation in the laboratory. Carefully designed and executed simulation may increase the generality of results. Because this strategy has been used with increasing frequency lately, a detailed discussion is in order.

Why Simulate? You may decide for a variety of reasons to simulate rather than conduct research in the real world. You may choose simulation because the behavior of interest could not be studied ethically in the real world. For example, Chapter 1 mentioned factors that control panic behavior. Re-creating a panic situation in order to study the ensuing behavior is unethical. If you were interested in studying how juries reach a decision, you could not eavesdrop on real juries. However, you could conduct a jury simulation study and analyze the deliberations of the simulated juries.

Often researchers choose to simulate for practical reasons. A simulation may be used because studying a behavior under its naturally occurring conditions is expensive and time consuming. By simulating in the laboratory, the researcher also gains the advantage of retaining control over variables while studying the behavior under relatively realistic conditions.

Designing a Simulation For a simulation to improve the generality of laboratory-based research, it must be properly designed. Observe the actual situation and study it carefully (Winkel & Sasanoff, 1970). Identify the crucial elements and then try to reproduce them in the laboratory. The more realistic the simulation, the greater are the chances that the results will be applicable to the simulated real-world phenomenon. As an example, suppose you were interested in studying the interpersonal relationships and dynamics that evolve in prisons. It might be difficult to conduct your study in an actual prison, so you might consider a simulation. In fact, Haney, Banks, and Zimbardo (1973) did just that.

In their now-famous Stanford prison study, Haney et al. (1973) constructed a prison in the basement of the psychology building at Stanford University. Participants in the study were randomly assigned to be either prisoners or prison guards. Those participants assigned to be prisoners were "arrested" by the police, fingerprinted, and incarcerated in the simulated prison. Treatment of the prisoner-participants was like that of actual prisoners: They were issued numbers and drab uniforms and were assigned to cells. Prison guards were issued uniforms, badges, and nightsticks. Their instructions were to maintain order within the simulated prison.

The behavior of the participants within the simulated prison was observed by a team of social psychologists. Behavior within the simulated prison was similar to (though less extreme than) behavior in a real prison. Guards developed rigid and sometimes demeaning rules, and prisoners banded together in a hunger strike. In fact, the simulation was so real for the participants that the experiment had to be discontinued after only a few days.

Realism Most researchers would agree that a simulation should be as realistic as possible (as was the case in the Stanford prison study). The physical reality created in the Stanford prison study probably helped participants become immersed in their roles. However, a simulation may not have to be highly realistic to adequately test a hypothesis. For example, many jury simulation studies do not re-create the physical setting of a courtroom. However, many of these studies are highly involving and compelling for the participants.

The importance of the "realism" of a simulation depends in part on the definition of *realism* that you adopt. Aronson and Carlsmith (1968) distinguish between two types of realism: mundane and experimental. The term *mundane realism* refers to the degree to which a simulation mirrors the real-world event. In contrast, *experimental realism* refers to the degree to which the simulation psychologically involves the participants in the experiment.

Simulation is an important issue in the area of social psychology and law. Many researchers have used simulation methods to study issues such as plea bargaining and jury decision making. A simulation in which a courtroom is realistically reconstructed in the laboratory could have high mundane realism. However, such high levels of mundane realism do not guarantee that the results of the study will be any more valid than those of the same study conducted in a more ordinary laboratory setting. Experimental realism is an important factor to be considered. An involving task in a laboratory with low mundane realism may produce more general results than a less involving task in a laboratory with high mundane realism.

A good illustration of the importance of experimental realism comes from a study by Wilson and Donnerstein (1977). These researchers report that a crucial factor in the applicability of simulated jury research findings is whether or not the participant believes that his or her decision will have real consequences. As an independent variable, Wilson and Donnerstein varied whether or not participants believed that their decisions would have consequences. They found that when participants believed that their judgments had consequences, the defendant's character (a variable previously shown in other research to be an important factor in the decision process) was no longer important.

Leading the participant to believe that his or her decision has consequences beyond the advancement of science increases experimental realism and thus increases the generality of the results. You may be able to increase the generality of your results when designing simulation studies by taking steps to increase not only mundane realism but also experimental realism.

To summarize, the laboratory approach to research has the advantage of allowing you to control variables and thus to isolate the effects of the variables under study. However, in gaining such control over variables, you lose a degree of generality of results. Using simulations that are high in experimental realism may improve the ability to generalize laboratory results in the real world.

The Field Setting

Field research is research conducted outside the laboratory in the participants' natural environment (the "field"). In this section, we briefly discuss conducting experiments in the field. However, most field research employs nonexperimental (correlational) methods such as naturalistic observation or survey designs. (We discuss these nonexperimental methods in Chapters 8 and 9.)

The Field Experiment A *field experiment* is an experiment conducted in the participant's natural environment. In a field experiment (as in a laboratory experiment), you manipulate independent variables and measure a dependent variable. You decide which variables to manipulate, how to manipulate them, and when to manipulate them. Essentially, the field experiment has all the qualities of the laboratory experiment except that the research is conducted in the real world rather than in the artificial laboratory setting.

As an example, consider an experiment conducted by Ute Gabriel and Rainer Banse (2006) to investigate whether gays and lesbians are the target of discrimination. Their measure of discrimination was whether gays and lesbians were helped less than heterosexuals. Residents of Berlin, Germany, were called between 6:00 P.M. and 9:00 P.M. over a 4-week period by a male or female researcher. The sex of the caller was communicated to participants by having the male researcher call himself Michael and the female researcher call herself Anna. Once a participant was on the telephone, the researcher asked the participant if the researcher's romantic partner was at home. Sexual orientation of the caller (researcher) was manipulated by having the caller ask for a same-sex (e.g., Michael asks for Peter) or opposite-sex partner (e.g., Anna asks for Peter). When the participant indicated that the caller had reached the wrong number, the researcher went on to explain that his or her car had broken down and that he or she did not want the romantic partner to worry. The participant was told further that the caller had no more money for another call and asked the participant to call his or her partner so that he or she would not worry. At this point, the caller gave the participant a number to call. The dependent variable was the number of participants in each experimental condition who made the call.

Gabriel and Banse (2006) found that homosexual callers were significantly less likely to receive help (67%) than heterosexual callers (83.5%). This difference was found for both male and female callers. They also found that male participants were significantly less likely to help homosexual callers than were female participants. Interestingly, Gabriel and Banse also report that male and female participants discriminated against lesbian callers at about the same rate. However, male participants discriminated against gay callers significantly more than female participants.

This field experiment has all the elements of a true experiment. Independent variables were manipulated (sex of caller and sexual orientation of caller) and a dependent variable was measured (whether the participant called the number provided by the caller). Hence, causal inferences about helping behavior can be made from the observations.

Advantages and Disadvantages of the Field Experiment As with the laboratory experiment, the field experiment has its advantages and disadvantages. Because the research is conducted in the real world, one important advantage is that the results can be easily generalized to the real world (i.e., high external validity). An important disadvantage is that you have little control over potential confounding variables (i.e., low internal validity). In the Gabriel and Banse (2006) field experiment, for example, the researchers could not control who would answer the telephone when the researcher called. Nor could they control how many others were present with the participant when called and what participants were doing when the call came in. Each of these variables could affect the reaction of a person asked to make a call for someone else. These extraneous variables can obscure or distort the effects of the independent variables manipulated in field experiments.

A Look Ahead

At this point, you have been introduced to the broad issues that you should consider when choosing a research design, the basic design options available to you, and the strengths and weaknesses of each choice. Before you are ready to conduct your first study, you also will need to know how to measure your variables; what methods of observation are available; how to conduct systematic, reliable, and objective observations; how to choose participants and deal with them ethically; how to minimize participant and experimenter biases; and many other details concerning specific research designs. In the next chapter, we consider how to go about making systematic, scientifically valid observations.

QUESTIONS TO PONDER

1. What is a simulation, and why would you use one?
2. How does the realism of a simulation relate to the validity of the results obtained from a simulation?
3. What are the defining features of laboratory and field research?
4. What are the relative advantages and disadvantages of laboratory and field research?

SUMMARY

Some of the most important decisions that you will make about your research concern its basic design and the setting in which it will be conducted. Research designs serve one or both of two major functions: (1) exploratory data collection and analysis (to identify new phenomena and relationships) and (2) hypothesis testing (to check the adequacy of proposed explanations). In the latter case, it is particularly important to distinguish causal from correlational relationships between variables. The relationship is causal if one variable directly influences the other.