Summary

As scientists of human behavior, psychologists have many research designs available to them, all of which aim to establish relationships between events and to fit these relationships into an orderly body of knowledge. Among the **quantitative designs** is the **experimental method,** which is the primary focus of this book. This method requires that a particular circumstance be manipulated and some aspect of behavior measured. From an experiment it is possible to say whether the manipulation of the circumstance caused any change found in the behavior.

Sometimes when an experimental approach cannot be used, it is necessary to use **correlational observations,** in which variables are observed and their relationships evaluated. The results of such a study cannot be used to establish causal relationships, because none of the variables is under the control of the investigator. Correlational observations are often carried out using a survey in the form of a questionnaire or interview. Correlational data can also be obtained by doing **archival research** with data contained in public or private records, such as census data or court records.

Some investigators are now doing research that employs **qualitative designs.** Qualitative researchers use descriptive data: written descriptions of people, including opinions and attitudes, and of events and environments. In **ethnography** the investigators use interviews and sometimes participatory observations to gather descriptive data. In one form of qualitative research they use **naturalistic observation,** in which data are gathered in realistic settings. A final qualitative design used when the potential number of observations is limited is the **case history,** in which detailed accounts of the events in a person's life or in a historical incident are described and analyzed.

2

How to Do Experiments

During its long history down to the middle of the nineteenth century, psychology was cultivated by able thinkers who did not realize their need of carefully observed facts. . . . Finally psychologists decided that they must follow the lead of physics, chemistry and physiology and transform psychology into an experimental science.

R. S. WOODWORTH (1940)

We must guard against . . . the drawing of a preconceived conclusion from experiments or observations which are so vaguely conditioned that a variety of inferences are as a matter of fact possible.

K. DUNLAP (1920)

In the first chapter we briefly discussed the experimental method. You will recall that the major advantage of doing this type of research is that it allows you to make causal statements—that a circumstance caused a change in behavior. Because this type of statement is precise, the rules required to support the statement are quite stringent. Most of these rules involve being able to account for all the circumstances that could vary.

By way of example, suppose that we were interested in the time it takes a person to press a button in response to a light of a particular intensity. At this point we have chosen a circumstance to manipulate—the intensity of a light—and a behavior to measure—the time to press a button. These two variables have formal names.

Variables

INDEPENDENT VARIABLES

The circumstance of major interest in an experiment, light intensity in our example, is called an **independent variable.** The best way to remember this name is to recall that the variable is independent of the participant's behavior.

As experimenters, we manipulate this variable—that is, choose two or more levels to present to the participant—and nothing the participant does can change the levels we have chosen. For example, if our independent variable is light intensity, we might select a high-intensity light and a low-intensity light as our two levels and observe behavior under both circumstances. Without at least two levels, we are not doing an experiment, but we are free to choose many more levels or to have more than one independent variable. In the later chapters I discuss ways of designing these more complex experiments.

DEPENDENT VARIABLES

Once we have chosen the independent variable, we will want to measure a participant's behavior in response to manipulations of that variable. We call the behavior we choose to measure the **dependent variable** because it is dependent on what the participant does.[1] In the reaction-time experiment, for example, our aim is to find out whether a relationship exists between light intensity and time to respond. Thus, our dependent variable is the time from when the light is turned on until the participant presses a button. Making a statement about the expected nature of the relationship is sometimes useful; such a statement is called a **hypothesis.** In the example, we might hypothesize that the more intense the light, the quicker the response will be. The outcome of the experiment will determine whether the hypothesis is supported and becomes part of the scientific body of knowledge or whether it is refuted.

I have discussed hypotheses in several different places in this book. In the next chapter we will consider how hypotheses can be deduced from theories and how they must be true if the theory is true. In Chapter 12 we will discuss the concept of a null hypothesis. As we will see, the null hypothesis is just a statistical statement saying that the independent variable has no effect on the dependent variable for a population. However, if you really believed there would be no effect in your experiment, you would probably not carry out that experiment. In actuality, you are usually predicting that the change in levels of the independent variable will cause a change in the dependent variable. This prediction is your real hypothesis. In fact, experimenters often go beyond this **nondirectional hypothesis** of simply predicting some change and state a **directional hypothesis** predicting the direction the dependent variable will change as the independent variable is manipulated.

In some cases hypotheses are not even based on theories, particularly when you simply wonder what would happen to a behavior if the independent variable were manipulated. In this case, the hypothesis is simply the answer to a question. How does crowding affect aggression? Does marking your first guess or thinking longer lead to better grades on multiple-choice tests? Are politicians who smile in their campaign posters more or less likely to win elections than those who don't? Hypothesized answers to questions such as these can also add to the scientific body of knowledge.

[1] I believe it is easier to remember the term this way, although the word *dependent* really refers to the behavior's being potentially dependent on the levels of independent variable.

CONTROL VARIABLES

So far, we have chosen one circumstance to manipulate—the independent variable. However, in some way, we will need to account for other circumstances in an experiment. One possibility is to control the other circumstances, thus making them **control variables.** We can control such circumstances by making sure that they do not vary from a single level. For example, in our reaction-time experiment, we might require constant lighting conditions in the room, only right-handed participants, a constant temperature, and so on. Ideally, all circumstances other than the independent variable would stay constant throughout an experiment. We would then know that any change in the dependent variable must be due to the changes we had brought about in the independent variable.

The concept of control is vital for experimentation and makes the experimental method distinct from the other forms of research that I discussed in the previous chapter. In your experiments, many of the variables will be set as control variables. As an experimenter, you will want to be sure that you have indeed achieved complete command of the control variables in your experiment. This is why psychologists go to considerable expense to build special environments in which sound, light, and temperature are controlled and to use special equipment to ensure that stimulus characteristics are consistent and that responses are carefully measured.

However, even though many variables in your experiments will be control variables, you should realize that, especially in psychology, not all variables will be assigned as control variables. First, the experimenter cannot control all the variables. It is impossible not only to control many genetic and environmental conditions but also to force cooperative attitudes, attentional states, metabolic rates, and many other situational factors on human participants.

Second, we really do not want to control all the variables in an experiment otherwise we would create a unique set of circumstances. If we could control all variables while manipulating the independent variable, the relationship established by the experiment would hold in only one case—when all variables were set at exactly the levels established for control. In other words, we could not *generalize* the experimental result to any other situation. As a rule of thumb, the more highly controlled the experiment, the less generally applicable the results.

Suppose, for example, that General Nosedive from the U.S. Air Force came to you and said: "Say, I understand you ran an experiment on reaction time. Tell me how intense I should make the fire-warning light in my fighter planes so that my pilots will respond within half a second." Having conducted a well-controlled experiment, you reply, "Sir, if you can guarantee that the pilot is a 19-year-old college sophomore with an IQ of 115, sitting in an air-conditioned, 10-foot-by-15-foot room, with no distracting sounds and nothing else to do, and if you always give a warning signal 1 second before the light comes on, then I might be able to give you an answer." You can probably imagine the general's reply. The moral of the story: if you want to generalize the results of your experiment do not control all the variables.

The generalizability of an experimental finding is also referred to as **external validity**—how well a causal relationship can be generalized across people, settings, and times. Cook and Campbell (1979) have defined several types of validity. The way they use this term, *validity* refers to whether drawing experimental conclusions about cause is justifiable. I will introduce other terms for validity at appropriate places in the book. Threats to external validity might occur if you use a limited sample, such as college sophomores, when you want to generalize to all humans of any age or intelligence (including, as in our example, Air Force pilots). Or you might have done a highly controlled laboratory experiment when you want to generalize to real-world work settings where it is noisy, hot, and crowded, and the workers are tired and unmotivated but have lots of practice. In general, the more tightly controlled your experiment—that is, the more circumstances you choose to make into control variables—the more likely it is to suffer from threats to external validity.

RANDOM VARIABLES

Having established that we do not want to control all the circumstances, what can we do with the remaining circumstances in our experiment? One possibility is to let them vary. In what way can we allow the circumstances to vary and still be sure that they will not bias our experiment? One alternative is to permit some of the circumstances to vary randomly. These variables are termed **random variables.**

The term *random* or *randomization* is used in several different ways in science. One use of the term is in the context of **random selection** of items from a population to form a representative sample. In this case, a population of items is available and some random process is used that makes the selection of any one item from that population as likely as the selection of any other item. Random selection is used to ensure external validity, that is, to ensure that the sample of items randomly selected from the population is generalizable to that population. So, if you wanted to generalize the results from an experiment to all people in the United States, ideally you would use a means of selection that was equivalent to putting the name of everybody in the country into an enormous hat and drawing out a sample of names. You could then say that you have randomly selected your sample and you could claim good external validity of your findings.

However, in this context the word *random* in the term random variable usually refers to **random assignment** of circumstances to the levels of the independent variable. Many of the circumstances in an experiment concern individual differences in the participants. Obviously, if we use the same participants for the various levels of the independent variable, we do not have to worry about individual differences. However, if we use different participants for each level of the independent variable, then we have to make sure that the characteristics of the participants assigned to each level do not bias our conclusions. For example, suppose that you want to determine the effects of TV violence on aggression in children. After you have randomly selected two hundred 6-year-old children as a sample from some larger population, you might then randomly assign them to two levels of the independent variable: viewing violent TV shows and viewing nonviolent TV shows. Perhaps you could flip a coin for each child and assign the child to the first group if a head occurred and to the second group if a tail occurred. Is it possible that most of the children in first group attend violent schools or eat lots of sugar or come from abusive homes, while few of those in the second group do? Yes, but if the selection was done in a truly random manner, it is statistically unlikely for such large samples to be biased.

Suppose that you let the children watch the violent or nonviolent TV shows at home. Is it possible that most of the children in one group have large-screen theater-system TVs at home, while most of those in the second group have small portable TVs? Again, it is possible but not probable; randomness makes this possibility highly unlikely.

There is no particular trick to random assignment or random selection. You can use any device that allows each item an equal chance of assignment or selection. As in the example, if you want to form two groups, you can flip a coin to form them.[2] If there are six groups, you can throw a die. If there are 33 groups, you can use 33 equal-sized slips of paper. Most mathematical handbooks and many statistics texts have random-number tables based on a process equivalent to drawing from 10,000 slips of paper. I have included one such table in the back of this book as Appendix C. Using any column or columns in a table of random numbers, you can assign each of your items a number and select the item when that number occurs. Just ignore the extra columns or numbers that are not on your list. If you happen to be a computer buff, you can use the computer to generate random numbers or events.[3]

If you have chosen to make a circumstance into a random variable, you must be sure that it varies in a truly random way, because not all events that appear random are really so. For instance, if you try to randomize conditions in an experiment by assigning events yourself, you have not randomized! Humans are notoriously bad at producing random events. If you assume that participants will show up for an experiment throughout the day or throughout the semester in a random order, you are wrong! People who are morning or afternoon volunteers or early-semester or late-semester volunteers have different characteristics. New experimenters commonly make mistakes in randomization. Don't you make them!

Perhaps most of the circumstances that become random variables in your experiment will be associated with participants and can be randomized by randomly assigning participants. However, other circumstances that are not associated with participants can sometimes be treated as random variables. Suppose in our TV violence experiment that the room in which the children

[2] Actually, most coins are slightly biased in favor of heads, but, unless your experiment has over 10,000 trials, don't worry about it.

[3] Computers are also less than perfect at generating random events, but they're much better than coins. For assigning events in an experiment, it doesn't make much difference which method you use.

itch TV is available either in the morning or in the afternoon. If you think ere is a reason why watching in the morning versus the afternoon may use differences in how aggressive children become independent of the ount of violence in the show, then you might want to randomly assign the olent-TV and the nonviolent-TV groups to morning and afternoon times. u certainly would not want one group to watch exclusively in the morning d the other in the afternoon.

There are other circumstances that also might affect the aggressiveness of ldren that you may consider as random variables even though you have t been able to make truly random assignments; for example, how stormy weather is or how much violence there is in the news on a particular day. ou have multiple sessions of your experiment, you would probably not be far off if you assume that these circumstances are randomly distributed oss the levels of your independent variable and that they will not system- cally bias your results.

As mentioned earlier, the major advantage of random selection is the eralizability of the results. Every time you choose to make a circumstance o a control variable, you can generalize the results to only that level of the iable. However, if you allow many levels of the circumstance to exist in opulation and then randomly choose a sample, you can generalize to entire population. The major advantage of random assignment is the nination of bias from the results. Thus, randomization can be a powerful erimental tool.

NDOMIZATION WITHIN CONSTRAINTS

some cases you may not want to make a circumstance into either a dom or a control variable. Actually, randomization and control define posite ends of a continuum. Falling between these two extremes are vari- degrees of **randomization within constraints.** In this case, you control part of the event assignments and randomize the other. Suppose that in reaction-time experiment we knew that practice could be an important iable.

If we presented all the low-intensity trials first, followed by all the high- nsity trials, we could be accused of biasing the experiment; any difference ween response times to low- versus high-intensity light might, in fact, be to short versus long practice. To avoid this problem, we could decide to trol the practice variable and give only one trial to each individual. Or we ld assign the low- and high-intensity trials randomly over, say, 12 trials flipping a coin and presenting a high-intensity light whenever a head urred and a low-intensity light whenever a tail occurred. However, this rnative might not be the most attractive one, because it could result in an dequate representation of high and low intensities. (For example, the flip- g of the coin might result in only three high-intensity trials and nine low- nsity trials.) To avoid this possibility, we could decide to have an equal nber of high- and low-intensity trials.

Thus, as a solution we establish a constraint on the assignment of trials (an equal number of each type of trial) and make a random assignment within this constraint. We might write the word *high* on six slips of paper and the word *low* on six and draw them out of a hat to determine the order of presentation. This procedure would fulfill the requirement that the conditions be randomly ordered across trials within the constraint that the two intensities be equally represented.

Other constraints, of course, are possible. We might want to avoid the possibility of too many trials at a particular intensity occurring early in the sequence. We could then **randomize within blocks,** with the block serving as our constraint. Using this alternative, we could choose three blocks of four trials each, ensuring that we randomly selected two high-intensity trials and two low-intensity trials within each block. To describe this procedure, we would say that we randomly assigned conditions to three blocks of four trials each, with the constraint of representing each intensity an equal number of times within each block.

You can legitimately use many such constraints as long as you specify them. However, the more constraints you specify, the less random your selection process is, and the less generalizable your results are.

CONFOUNDING VARIABLES

If we designed our experiment perfectly so that we have chosen an independent variable to manipulate and a dependent variable to measure and made the rest of the circumstances into control variables, random variables, or variables randomized within constraints, then we would not have to worry about the variable I am about to discuss. However, not every experiment is designed perfectly, and in many real-world settings, designing a perfect experiment is impossible. In this case we need to know when a confounding variable rears its ugly head. Any circumstance that changes systematically as the experimenter manipulates the independent variable is a **confounding variable.**

Suppose, for example, that we used three different light intensities in our reaction-time experiment: a low-intensity light for the first 20 trials, a medium-intensity light for the next 20, and a high-intensity light for the last 20. If we reported, "People respond more quickly the more intense the light," someone else could say, "No, people respond more quickly after practice." In fact, we could both be correct, or either one of us could be incorrect! The problem is that we have unintentionally *confounded* the experiment with a variable that changes systematically with the independent variable.

An experimenter can record the most sophisticated measurements, do the finest statistical test, and write up the results with the style of Hemingway, yet a confounding variable can make the whole effort worthless. A feud between Coca-Cola and PepsiCo illustrates the type of confusion that this variable can cause ("Coke-Pepsi Slugfest," 1976). PepsiCo pitted its cola against Coke in a drinkers' test in which tasters who said they were Coke drinkers drank Coke from a glass marked Q and Pepsi from a glass marked M. More than half the tasters reportedly chose the glass containing Pepsi as their favorite. Coca-Cola officials countered by conducting their own preference test—not of colas but of letters. They claimed that more people chose glass M over glass Q not because they preferred the cola in glass M but because they liked the letter M better than they liked Q. This hypothesis was supported when most people tested still claimed to prefer the drink in the M glass when *both* glasses contained Coke.

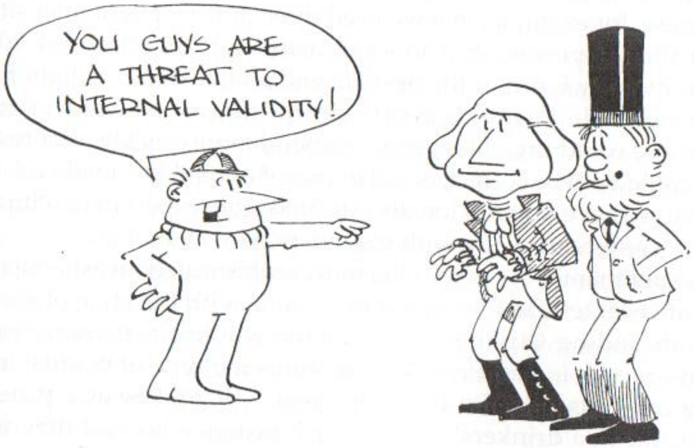
In this example, the letters were apparently a confounding variable. Because they varied systematically with the colas in the original test, the experimenters could not distinguish the tasters' preference for the colas from their preference for the letters.

I mentioned earlier in this chapter that Cook and Campbell (1979) have identified several types of validity. Another type is **internal validity,** which refers to whether the manipulated change in the independent variable caused the change in the dependent variable or whether something else caused the change. If the independent variable didn't cause the change, then a confounding variable must have. So, if we want to avoid confounding variables in our experiments, we need to understand the various possible threats to internal validity. There is no more important task for you as an experimenter than being able to recognize and, if possible, avoid the threats to internal validity that may introduce confounding variables into your experiments.

■ Threats to Internal Validity

HISTORY

In laboratory experiments, one can usually collect data at all levels of the independent variable over a relatively short time span. In this case, any change in the dependent variable is unlikely to have been due to **history**—some



event that takes place between the testing of the levels of the independent variables.

For example, suppose that you wanted to find out whether using computer-generated visuals rather than traditional hand-drawn overhead transparencies in a large introductory psychology class improves grades. Further suppose that a particular professor teaches this introductory course only once a year. For practical reasons you decide to ask this professor to use the computer-generated visuals this year and compare the grades of this class to those of the previous year's class. If you find that the overall grades are better for this year's class, you might be correct in attributing the improvement to the use of the computer-generated visuals. However, some historical event could have caused the change. For example, the school could have tightened admission standards, thereby changing the academic quality of students in the class. Or perhaps the college of engineering decided to require all senior engineering students to take the class, again changing the class composition. Or perhaps the world has undergone an increase in interest for the subject matter being taught, similar to what occurred in computer science courses after personal computers appeared. Or perhaps, at a local level, a fraternity has acquired a copy of last year's test and has made it available to certain students in the class. To have confidence in the conclusion that the change in grades between the classes was due to the use of computer-generated visuals, you must rule out these historical events, as well as any others that might threaten the internal validity of the conclusion.

MATURATION

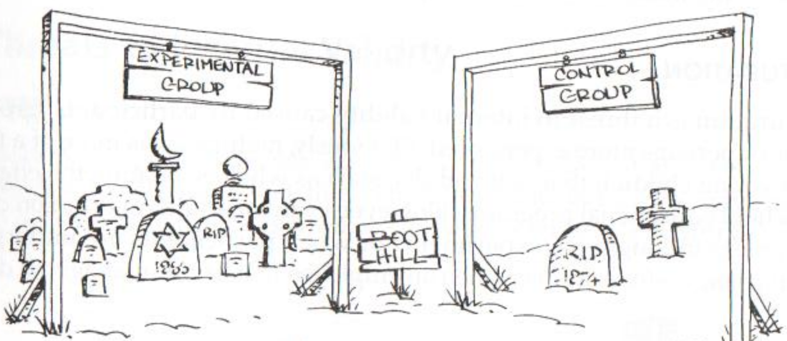
Maturation is a threat to internal validity caused by participants' growing older or perhaps more experienced. Obviously, maturation is more of a threat with young children than with adults, such as when evaluating the effects of preschool educational programs. However, even for adults, maturation can be a problem in long-term experiments or when participants are undergoing rapid change—for example, when an employee first begins managerial duties.

SELECTION

Selection can be a threat whenever experimenters cannot assign participants randomly, particularly when, for practical reasons, participants essentially assign themselves to conditions. If, in the previous example, the classes chosen for comparison were fall and spring classes, selection could be a problem. My fall introductory psychology class contains many first-year students, many of whom are psychology majors. The spring class has many more engineering students who have put off taking this required course until their senior year. Do you think there might be differences between these classes besides the use of computer-generated visuals? Experimenters who use college students as participants are familiar with the potential differences between early-semester volunteers[4] and late-semester volunteers. In general the early-semester volunteers are more eager and motivated and are probably better students or at least better at planning their time. The worst kinds of selection threats are those that are directly linked to the independent variable. For instance, suppose that you want to evaluate a new industrial training program. You let workers volunteer to take the new program and then after completing the program you compare the performance of those workers to the performance of workers who did not volunteer for the program. Do you think there might be a difference in the workers who self-selected to be in the two groups? What about the difference between the recovery rates of people who choose a new type of therapy and the rates of those who refuse the therapy?

MORTALITY

Participants dropping out of an experiment, **mortality,**[5] can also be a threat to internal validity. Fortunately, in most experiments, these participants die only with respect to their life in the experiment, not with respect to life in general. *Overall* mortality is not really a problem; *differential* mortality is a problem. This occurs when more or different kinds of participants drop out of the groups assigned to various levels of the independent variable. For example, suppose that a company decides to try a new training program to inoculate newly promoted middle managers against socially stressful situations.



THE THREAT OF DIFFERENTIAL MORTALITY

The company randomly chooses half of its new managers for a 1-hour-per-day exposure to simulated personal confrontation with employees. The other managers are not exposed to this training program. The number of stress-related health complaints in the two groups is counted for a period of

[4] A *volunteer* in this case is a little like a volunteer in the military. Although some of the students who crowd around experimental sign-up sheets would volunteer even if such service were not a course requirement, most actually volunteer to do this in place of some other requirement, such as writing a paper.
[5] *Mortality* is the term used by Cook and Campbell (1979). Some experimenters also refer to this as *attrition*.

five years after the training. The company finds that the stress-inoculated group has reported fewer such complaints and concludes that the program was a success. Was it? Among the questions that you should ask is: How many managers dropped out of each group during the training program?[6] Conversely, the training may have sensitized the managers to be more aware of stress-related health problems. It is likely not only that more managers would have dropped out of the stress group but also that these managers would have been the most sensitive to stress. The success of the training group might have little to do with the inoculation procedure but might be due entirely to the fact that mortality changed the characteristics of the groups.

TESTING

The act of testing can change behavior independently of any other manipulation. Testing can be a threat to internal validity when a pretest or multiple-test design is used. Suppose that you are interested in whether a new advertising campaign would increase the public's awareness of your company's brand of shaving cream. You pick a large random sample of consumers and send them a questionnaire. You ask a number of questions about various brands of shaving cream and the commercials associated with the brands. Three months later, after launching a new series of commercials touting your brand, you again send the questionnaire to the same people and discover that they are now much more familiar with your brand of shaving cream. You declare the advertising campaign a success. Are you right?

One problem with the conclusion that the campaign caused a change in awareness is that the pretest itself may have caused the change in awareness. The pretest may have sensitized this particular group of people to notice shaving cream brands in general. During the following three months, they may have watched all the shaving cream commercials more closely, and now they are able to tell you more about each of the brands regardless of the new advertising campaign.

In addition to sensitizing participants, testing can also inform the participants of the experimenter's topic of interest or even the experimental hypothesis. A pretest can also provide information, increasing the participants' knowledge of a topic so that scores on a posttest will be higher, independent of any experimental manipulation.

STATISTICAL REGRESSION

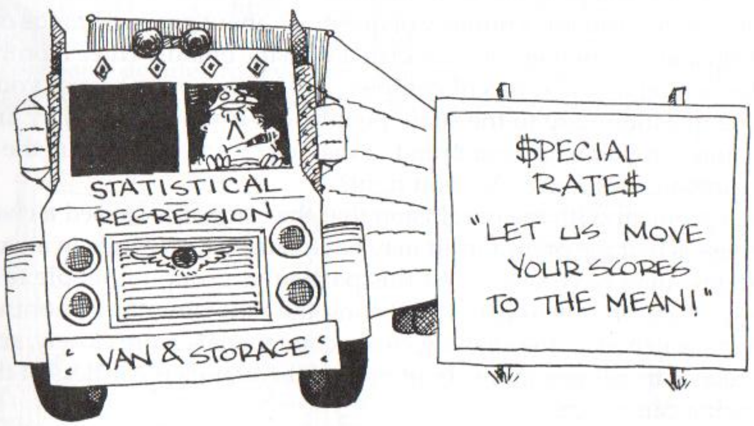
Perhaps the most subtle threat to internal validity is **statistical regression.** This term refers to the fact that when experimenters choose participants on the basis of their having scored very high or very low on a particular test,

[6] In addition to the threat to internal validity of mortality, which I am emphasizing here, you should be able to find other potential threats. For example, the training program might harbor demand characteristics (see Chapter 4) that bias these managers against reporting stress-related health problems.

their scores tend to move toward the mean on a second test. It is not immediately obvious why regression toward the mean should occur. Perhaps an example will help.

Suppose that you have devised a program that you claim will increase the IQ scores for preschool children who have been classified as having mild retardation (IQ of 53 to 68). You give an IQ test and choose 30 children who score within the mild retardation range. After one year in your program, you give the children the test again. You discover that the mean IQ of the group has increased by seven points and that this change is statistically significant. You declare your program to be a success. Is it?[7]

How could statistical regression have caused or contributed to this result? Imagine that the IQ pretest is composed of two separate components: a "true" IQ that a perfect test would measure and "error." The perfect test yields exactly the same score for a particular child every time you give it. If you could use such a test, statistical regression would pose no problems. But, alas, the IQ that you measure also has an error component.



This error may be due to a number of unpredictable variables. For instance, the child might have been lucky and have guessed the correct answers to several items on the pretest, or might have been unlucky and have guessed fewer correct answers than chance would predict. Or perhaps the child got up on the wrong side of the crib that morning and had a difficult time concentrating on the pretest. Or perhaps the examiner was feeling particularly grouchy that morning and failed to establish good rapport with the child. Because we cannot predict the size or direction of this error component for any particular score,[8] we must treat error as if someone were drawing a random number out of a hat and adding or subtracting it from the true score.

When you chose the group with mild retardation on the basis of a low pretest score, you probably chose many more children who had error working against them than children who had error artificially inflating their true score. That is, the true scores of this group were, on the average, not as low as the ones they received on the pretest. Because you chose only children with low scores, you biased the group toward those with error working against them. However, on the retest one year later, we would expect a less biased error. We would expect as many errors to increase the true scores as to decrease them. There is still an error component, but now it is not biasing the measured score away from the true score.

If you are not yet convinced, try a little demonstration. Pick any true score, say, 100. Write the numbers from –10 to +10 on equal-sized slips of paper and put them into a container. Draw a number from the container, add it or subtract it from 100, write down the result, and replace the number. After doing this 30 times, take the lowest five numbers and figure the mean (add the numbers and divide by 5). Now, follow the same procedure, drawing just five numbers and figuring the mean. Is the first mean lower than the second mean? You have just demonstrated statistical regression.

INTERACTIONS WITH SELECTION

Finally, validity threats such as maturation and history may have **interactions with selection.** As an example of the possible interaction of selection with history, consider the following study. Stanley Coren and his colleagues studied archival records and found that the distributions of right- and left-handers in age-groups ranging from 10-year-olds to 80-year-olds were very different (Coren & Halpern, 1991; Porac & Coren, 1981). From a high of 15% left-handers among the 10-year-olds, the percentage declined until there was 0% in the 80-year-old group. They concluded that left-handers had a "decreased survival fitness" that caused them to die at earlier ages. Obviously, many left-handers, mothers of left-handers, and husbands of left-handers were concerned about this conclusion. However, Lauren Harris (1993a) disputed the conclusions and presented evidence that the interaction of selection and history could have caused the change in percentages. Eighty years ago, people attached considerable stigma to being left-handed. So parents and teachers strongly encouraged children to be right-handed, forcing them to eat, write, and do other tasks with their right hands. In other words, selection by means of social pressure occurred for right-handedness. But this selection changed with history. Over the years, being left-handed became more acceptable, and

[7] At this point, you should be able to identify a number of potential threats to internal validity other than regression. The problem with maturity over a one-year period for preschool children is obvious. Testing could be a problem as well. The IQ pretest was probably the first test of any kind that these children had taken. They might have learned something in general about taking tests. They might also have remembered specific items from the pretest and learned the answers over the year.

[8] If you are using a standardized test, you might be able to get an idea of the general magnitude of the error component for the test—a number that characterizes the reliability of the test. The lower this number, the more we must be concerned about the effects of statistical regression.

parents and teachers pressured fewer truly left-handed children to become right-handers. So Harris argued that fewer left-handers are in the older groups not because most left-handers have died but because there never were many to begin with. The argument is not yet settled (see Halpern & Coren, 1993; Harris, 1993b), but the case of the disappearing left-handers offers an interesting example of the possible interaction of selection with history.

I hope this discussion of threats to internal validity will help you in your search for confounding variables. Whenever you are planning an experiment, going over each of these threats might be helpful to make sure that none of them is a problem for your experiment. In some cases you may have potential threats that are difficult or impossible to eliminate, in which case it might be possible for you to use a quasi-experimental design, many of which I have discussed in Chapter 10.

■ Summary of the Experimental Method

Now that you are familiar with the use of the experimental method, let's try to fit all the terms we have learned into a schematic framework. In Figure 2-1 I summarize the experimental model. On the left, I have listed all the circumstances that may affect behavior. On the right, I have listed all the potentially measurable behaviors. At the top, on the left, I have chosen one of the circumstances for manipulation, the independent variable. On the right, I have selected one of the behaviors for measurement, the dependent variable. The arrow indicates that we are interested in whether the independent variable causes a change in the dependent variable. Although we can ignore the other behaviors, we need to make sure we can account for all the circumstances. In the figure, I have partitioned these circumstances into control variables, random variables, variables randomized within constraints, and confounding variables. While partitioning the variables, we should keep in mind that a decision to control increases the precision of the results (internal validity) but decreases their generality (external validity). On the other end of the continuum, a decision to randomize decreases the precision but increases the generality.

As a final example to illustrate the types of variables that go into an experiment, I will describe an experiment that two colleagues and I conducted (Grobe, Pettibone, & Martin, 1973) and list some of the variables in a figure similar to Figure 2-1. We were interested in whether an instructor's lecture pace makes a difference in how attentive the students are. I was teaching introductory psychology to a class of 200 students at that time, so I had the dubious honor of trying to lecture to this class at different speeds. As the independent variable, then, we chose three different lecture paces. I attempted to lecture at each pace for at least five minutes during each lecture. We tape-recorded this portion of each lecture and counted the number of syllables per minute to make sure that my pace was within the allowable range of error. In Figure 2-2 you will see lecture pace listed as the independent variable. We could have

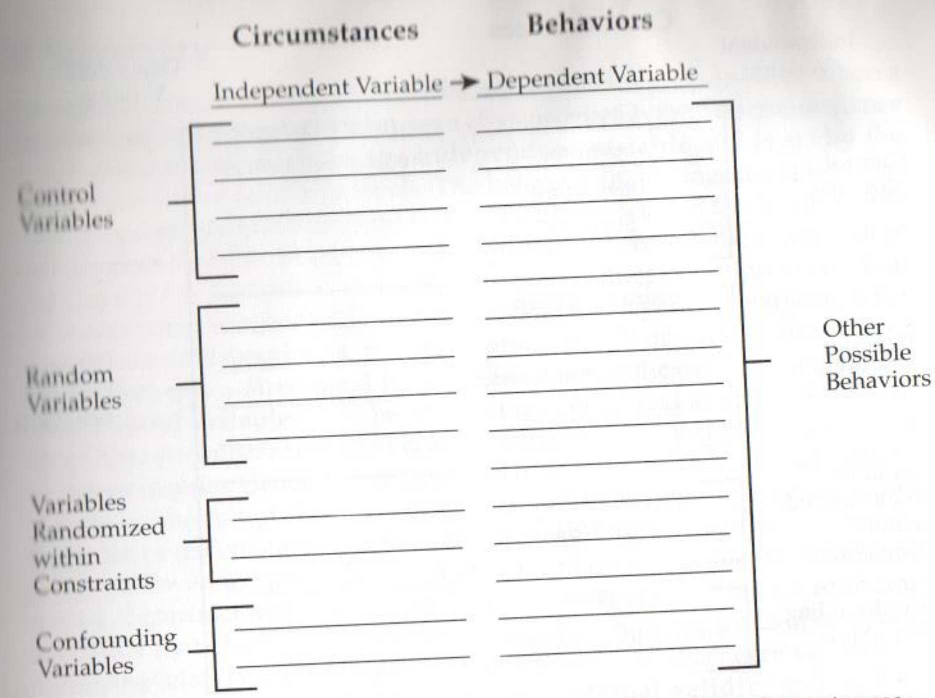


FIGURE 2-1 A diagram representing an experiment. One of the circumstances has been chosen as the independent variable. The others have been partitioned into control variables, random variables, variables randomized within constraints, and confounding variables. One of the behaviors has also been chosen as a dependent variable.

measured the students' attentiveness many ways: we could have videotaped the students and judges could have inferred their attentiveness, students could have filled out a questionnaire indicating how attentive they had been to each lecture, and so forth. Thus, we could have chosen many behaviors as dependent variables. To get a reliable quantitative measure, we recorded the background noise level in the room and inferred that when students were quietest, they were most attentive. So in Figure 2-2 you will find noise level listed on the behavior side as the dependent variable.

Many variables became control variables and did not change throughout the experiment: the classroom, the instructor, the time of day I gave the lecture, the students in the class, and so on. Some of these are listed as control variables in Figure 2-2. We allowed other variables to vary in an uncontrolled and (we hoped) random way, such as how much sleep I got the night before, the weather outside, the success of the football team each week, how many people in the class had colds (and coughed out loud), and many others. Some of these variables are listed as random variables in the figure. We randomized one variable within constraints. Because we were afraid that the day of the week might affect attentiveness, we did not want to have all the slow-paced

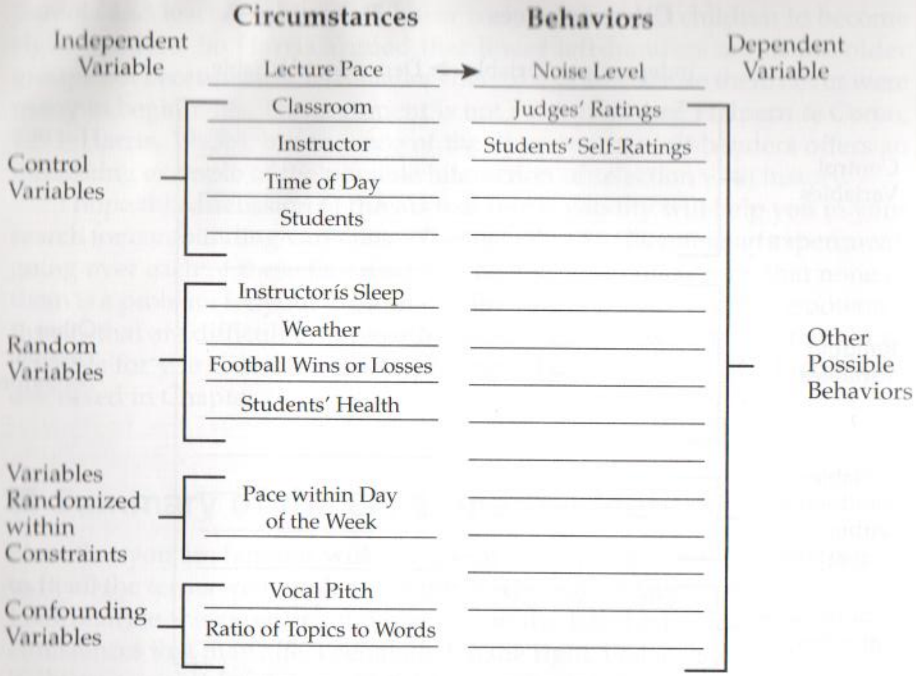


FIGURE 2-2 Variables from the lecture-pace experiment partitioned into an independent variable, control variables, random variables, a variable randomized within constraint, confounding variables, and a dependent variable

lectures on Mondays, the medium ones on Wednesdays, and the fast ones on Fridays. Therefore, we randomized the day of the week I would use each pace within the constraint that each day I gave each pace the same number of times.

Finally, although we tried to minimize confounding variables, we knew that, as in many applied experiments, some would occur. One was, undoubtedly, the average pitch of my voice. I am not a machine, so—as with most people—the faster I talk, the higher my voice becomes. I am sure that vocal pitch was confounded with lecture pace. In addition, as long as the length of a lecture remained constant, as I talked faster I could either say more words about a particular topic or say the same number of words and vary the number of topics I covered. I tried to do the former, so the ratio of topics to words was necessarily confounded with lecture pace. I have also listed these two confounding variables in Figure 2-2. I hope that this example illustrates how variables can be partitioned into the various types of circumstances and behaviors.[9]

[9] A reader commented that she wanted to know the outcome of this experiment. Briefly, we found that lecture pace did affect attentiveness. Fortunately, ambient noise levels were lowest for my medium pace. The noise levels were highest for my fast pace. So, we inferred that a medium pace is best, and it is better to err on the side of going too slow than too fast.

Summary

The experimental method allows causal statements to be made—that when a circumstance is manipulated it causes a change in behavior. The circumstance that is manipulated is called the **independent variable** and is set by the experimenter to at least two levels. The behavior that is measured is called the **dependent variable** because it may be dependent on the levels of the independent variable. The predicted relationship between the independent and dependent variable is call a **hypothesis**. If the prediction is just that the independent variable will cause a change in the dependent variable, it is a **nondirectional hypothesis**, but if the prediction is about the direction of change then it is a **directional hypothesis**. Some of the other circumstances called **control variables** may be set at a particular level and not allowed to vary. Other circumstances called **random variables** may be allowed to vary in a random manner. Generally, random variables improve the **external validity** of an experiment, and allow it to be generalized to other people, settings, and times. Some circumstances called **variables randomized within constraints** may be allowed to vary randomly but within limits imposed by the experimenter. Experimenters should attempt to eliminate or minimize **confounding variables** that change systematically with the independent variable and distort the relationship between the independent and dependent variables.

Confounding variables can cause low **internal validity** and make it difficult to say that only the independent variable caused a change in the dependent variable. Threats to internal validity include **history,** the occurrence of an uncontrolled event during the experiment; **maturation,** the change in age or experience of individuals during experimentation; **selection,** the biased assignment of individuals to groups; **mortality,** the nonrandom loss of individuals from groups; **testing,** the change in participants due to the testing process; **statistical regression,** the movement of scores toward the mean for groups selected on the basis of extreme scores; and **interactions with selection,** the differential effect of a threat on nonequivalent groups.