shau_c01.qxd 6/29/99 12:09 PM Page 1

PART I

General Issues

1

Introduction

Chapter Outline

Psychology as a Science

- The scope of scientific psychology has increased dramatically in the 100-plus years since its inception.
- The American Psychological Association (APA) and the American Psychological Society (APS) promote the science of psychology.

There is no record of who first observed behavior carefully and systematically, who conducted the first public opinion survey, or even of who performed the first psychology experiment. We don't really know exactly when psychology first became an independent discipline. It emerged gradually, with its roots in the thinking of Aristotle (the "father" of all psychology; Keller, 1937), in the writings of later philosophers such as Descartes and Locke, and, even later, in the work of early 19th-century physiologists and physicists.

The date usually taken to mark psychology's official beginning is 1879. In that year Wilhelm Wundt established a formal psychology laboratory in Leipzig, Germany. Wundt, like many scientists of his time, had a doctoral degree in medicine. He even published for a while in the fields of anatomy and physiology. These experiences gave Wundt a basis for his ideas about a "physiological psychology" that led to his scientific approach to psychology. Dozens of researchers from around the world came to his psychology laboratory to learn about the "new" discipline. As a consequence, from Wundt's laboratory a scientific psychology spread quickly to the intellectual centers of the world. According to one noted historian, Wilhelm Wundt was the first person "who without reservation is properly called a psychologist" (Boring, 1950, p. 316).

Whatever exact date one accepts for the beginning of a scientific psychology, we can see that the enterprise is little more than 100 years old. Near the end of the 19th century, in 1892, the American Psychological Association (APA) was formed, with G. Stanley Hall as its first president. Hall had been a student at

Wilhelm Wundt
(1832–1920)

William James
(1842–1910)

Harvard of the well-known American philosopher and psychologist William James. In 1878, James bestowed upon Hall the first Ph.D. in psychology given in the United States. After receiving his degree, Hall left the United States to study with Wundt in Germany (Boring, 1950). Still later, in 1887, Hall founded the first psychology journal in the United States, the *American Journal of Psychology*. The APA had only a few dozen members in its first year; 100 years later, in 1992, when the APA celebrated its 100th birthday, there were approximately 70,000 members (see American Psychological Association, 1991).

Psychology has changed significantly since its beginnings. Wundt and his colleagues were primarily interested in questions of sensation and perception—for instance, visual illusions and imagery. Experiments on reaction time also were conducted with the goal of measuring the time necessary for various cognitive processes, such as those involved in making choices (Boring, 1950). Psychologists continue to be interested in sensation and perception, and Wundt's procedures form the basis of some of the experimental methods in use today. However, research on sensation and perception represents but a fraction of the research undertaken in contemporary psychology. Today psychologists are interested in a myriad of topics, including those in such general areas as clinical, social, industrial, counseling, physiological, cognitive, educational, and developmental psychology.

Psychology has also not developed strictly as a laboratory science. Although laboratory investigation remains at the heart of psychological inquiry, psychologists and other behavioral scientists do research in schools, clinics, businesses, hospitals, and other nonlaboratory settings, including the Internet (see, for example, Smith & Leigh, 1997). Many research psychologists, for instance, are engaged in program evaluation, a type of research in which the impact of large-scale interventions on groups or institutions is evaluated. Program evaluators confront questions such as whether a new management–labor relations program will lead to increased employee morale and company productivity, or whether the nationwide Head Start program has significantly raised the IQ of hundreds of thousands of preschoolers (see Chapter 10).

Promotion of psychological research is a concern of the APA as well as the American Psychological Society (APS). Formed in 1988 to emphasize mainly scientific issues in psychology, APS has grown at a phenomenal rate. When APS celebrated its 10th anniversary in 1998, there were more than 15,000 members (see Ruksznis, 1998). Although many psychologists choose to join either APA or APS, many join both. APA and APS both sponsor annual conventions, which psychologists attend to learn about the most recent developments in their fields; each organization also publishes scientific journals in order to communicate the latest research findings to its members and to society in general; and both organizations encourage student affiliation, which provides educational and research opportunities for both undergraduate and graduate psychology students. By affiliating with APA and APS, students can subscribe to major psychology journals at a relatively low cost as well as become involved at an early stage in a career in psychology. Information about joining APA and APS as a regular member or as a student affiliate can be obtained by writing to the respective head offices, or by consulting their World Wide Web pages on the Internet at:

(APA) http://www.apa.org
(APS) http://www.psychologicalscience.org

Both the APA and APS web sites provide news about important recent psychological research findings, information about psychology publications, and links to many psychology organizations. Take a look!

The Scientific Method

- Psychologists use the scientific method to gain knowledge about human and animal behavior.
- The scientific method differs from nonscientific ("everyday") approaches to gaining knowledge.

There is one way in which psychology has not changed in the 100 years or so of its existence: The **scientific method** is still emphasized as the basis for investigation. The founding of Wundt's laboratory marked the beginning of the formal application of the scientific method to problems in psychology. This method is not identified with particular kinds of equipment, nor is it associated exclusively with specific research procedures. The scientific method is something abstract. It is an approach to knowledge that is best described by distinguishing it from what might be called *nonscientific* or "everyday" approaches to knowledge.

Scientific and Nonscientific Approaches to Knowledge

- The scientific method is empirical and requires systematic, controlled observation.
- To achieve control in a research situation, researchers manipulate independent variables or select levels of individual differences variables to determine their effect on behavior.
- Dependent variables are measures of behavior used to assess the effects of independent variables.
- Scientific reporting is unbiased and objective; clear communication of concepts occurs when operational definitions are used.
- Scientific instruments are accurate and precise; physical and psychological measurement should be valid and reliable.
- A hypothesis is a tentative explanation for a phenomenon; testable hypotheses have clearly defined concepts (operational definitions), are not circular, and refer to concepts that can be observed.
- Scientists adopt a skeptical attitude and are cautious about accepting explanations until sufficient empirical evidence is obtained.

TABLE 1.1 CHARACTERISTICS OF SCIENTIFIC AND NONSCIENTIFIC (EVERYDAY) APPROACHES TO KNOWLEDGE*

	Nonscientific (everyday)	Scientific
General Approach:	Intuitive	Empirical
Observation:	Casual, uncontrolled	Systematic, controlled
Reporting:	Biased, subjective	Unbiased, objective
Concepts:	Ambiguous, with surplus meanings	Clear definitions, operational specificity
Instruments:	Inaccurate, imprecise	Accurate, precise
Measurements:	Not valid or reliable	Valid and reliable
Hypotheses:	Untestable	Testable
Attitude:	Uncritical, accepting	Critical, skeptical

*Based in part on distinctions suggested by Marx (1963).

Several major differences between a scientific and a nonscientific approach to knowledge are outlined in Table 1.1. Collectively, the characteristics listed under "Scientific" define what is called the scientific method. The distinctions made in Table 1.1 between a scientific and a nonscientific approach are intended to highlight differences that frequently exist between an informal and casual approach that often characterizes our everyday thinking and thinking that is characteristic of the scientist's approach to knowledge. These distinctions are briefly summarized as follows.

General Approach Many everyday judgments are made intuitively. This usually means that we act on the basis of what "feels right" or what "seems reasonable." Intuition is not based on a formal decision process, such as that of deductive logic, nor is it based on information that was taught to us or that we acquired through direct experience. The many everyday inferences and conclusions reached intuitively are the product of insight and of what we quickly perceive as true.

Intuition is a valuable cognitive process. We frequently have little to go on other than what intuition suggests is the right answer or the proper course to follow. Intuition can also help us make decisions in situations that we have not encountered before. Consider, for example, what intuition might suggest as the answer to a question based on the following situation:

> Assume that on two different occasions you are the victim of a street crime. Let us say that you are walking down the street and someone tries to grab your wallet or purse. The thief begins to wrestle with you in order to get it. You find yourself in need of assistance. On one occasion there is a crowd of bystanders who witness the attempted robbery. On the other occasion there is only one person who sees the event. The question is: In which situation would you be more likely to receive help, when many people are present or when only one person is present?

Intuition would suggest that the more people present, the greater the chances that at least one person would help. Surprisingly, social psychologists have determined that just the opposite is true. A bystander is more likely to act in an emergency when alone than when a group of people is present (Latané & Darley, 1970; see also Chapter 3).

Our intuition about what is true does not always agree with what is actually true because we fail to recognize that our perceptions may be distorted by what are called cognitive biases or because we neglect to weigh available evidence appropriately (Kahneman & Tversky, 1973; Tversky & Kahneman, 1974). One such cognitive bias, for instance, is called *illusory correlation*. This is our tendency to perceive a relationship between events when none exists. To illustrate this bias, Ward and Jenkins (1965) showed people the results of a hypothetical 50-day cloud-seeding experiment. For each of the 50 days, the research participants were told whether cloud seeding had been done and whether it had rained on that day. Ward and Jenkins constructed the results such that there was actually no relationship between cloud seeding and the likelihood of rain—rain was equally likely on days when seeding had and had not been done. Nonetheless, people were convinced that the evidence supported their intuitive supposition that cloud seeding and rain varied together. One possible basis for the illusory-correlation bias is that we are more likely to notice events that are consistent with our beliefs than events that violate our beliefs. Thus, in Ward and Jenkins' study, people may have been more likely to notice and remember the days on which cloud seeding was followed by rain than the days on which clouds were seeded in vain because they believe cloud seeding produces rain.

The scientific approach to knowledge is empirical rather than intuitive. An **empirical approach** emphasizes direct observation and experimentation as a way of answering questions. This does not mean that intuition plays no role in science. Any scientist can probably recount tales of obtaining empirical results that intuition had suggested would emerge. On the other hand, the same scientist is also likely to have come up with just as many findings that were counterintuitive. Research at first may be guided by what the scientist's intuition suggests is the proper direction to take. Eventually, however, the scientist strives to be guided by what direct observation and experimentation reveal to be true.

Observation We can learn a great deal about behavior by simply observing the actions of others. However, everyday observations are not always made carefully or systematically. Most people do not attempt to control or eliminate factors that might influence the events that they are observing. As a consequence, erroneous conclusions are often drawn. Consider, for instance, the classic case of Clever Hans. Hans was a horse who was said by his owner, a German mathematics teacher, to have amazing talents. Hans could count, do simple addition and subtraction (even involving fractions), read German, answer simple questions ("What is the lady holding in her hands?"), and give the date, and tell time (Watson, 1914/1967). Hans answered questions by tapping with his forefoot or by pointing with his nose at different alternatives shown to him. His owner considered Hans to be truly intelligent and denied using any tricks to guide his horse's behavior. And, in fact, Clever Hans was clever even when the questioner was someone other than his owner.

Newspapers carried accounts of Hans's performances, and hundreds of people came to view this amazing horse (Figure 1.1). In 1904 a scientific commission was established with the goal of discovering the basis for Hans's abilities.

FIGURE 1.1 Top: Clever Hans performing before onlookers. Bottom: Hans being tested under more controlled conditions when Hans could not see the questioner.

The scientists found that Hans was no longer clever if either of two circumstances existed. First, Hans did not know the answers to questions if the questioner also did not know the answers. Second, Hans was not very clever if he could not see his questioner. It was discovered that Hans was responding to very slight movements of the questioner. A slight bending forward by the questioner would start Hans tapping, and any movement upward or backward would cause Hans to stop tapping. The commission demonstrated that questioners were unintentionally cuing Hans in this way.

This famous account of Clever Hans illustrates the fact that scientific observation (unlike casual observation) is systematic and controlled. Indeed, it has been suggested that **control** is the essential ingredient of science, distinguishing it from nonscientific procedures (Boring, 1954; Marx, 1963). In the case of Clever Hans, investigators exercised control by manipulating, *one at a time,* conditions such as whether the questioner knew the answer to the questions asked and whether Hans could see the questioner (see Figure 1.1). By exercising control, taking care to investigate the effect of various factors one by one, the scientist seeks to gain a clearer picture of the factors that actually produce a phenomenon.

The factors that the researcher controls or manipulates in order to determine their effect on behavior are called the **independent variables.** In the simplest of studies, the independent variable has two levels. These two levels often represent the presence and the absence of some treatment, respectively. The condition in which the treatment is present is commonly called the *experimental condition*; the condition in which the treatment is absent is called the *control condition*. If we wanted to study the effect of drinking alcohol on the ability to process complex information quickly and accurately, for example, the independent variable would be the presence or absence of alcohol in a drink that participants were given. Participants in the experimental condition would receive alcohol, while participants in the control condition would receive the same drink without alcohol.

The levels of the independent variable do not always represent the presence and the absence of some treatment; moreover, an independent variable may have more than two levels. What is critical is that the levels differ with respect to the variable of interest. For example, Heath and Davidson (1988) recruited female college students to participate in a study designed to aid in the development of a rape-prevention pamphlet. When volunteers appeared in the laboratory, they were asked to review one of three packages of materials that they were told was being considered for possible inclusion in the pamphlet. The three packages of materials varied in the manner in which rape was described. Specifically, the materials to be reviewed presented rape as being very controllable, somewhat controllable, or not at all controllable. After reviewing the materials, the women were asked a series of questions regarding their own perceptions of the risk of rape. The independent variable in this study was the degree of control emphasized in the pamphlet materials; it had three levels (high, medium, low).

Sometimes the levels of the independent variable are selected by a researcher rather than manipulated. This is typically the case when an **individual differences variable** serves as the independent variable. An individual differences variable is a characteristic or trait that varies consistently across individuals. For example, a researcher may be interested in how depression affects perceptions of other people (e.g., Albright & Henderson, 1995). Do depressed individuals, for instance, perceive others more or less accurately than individuals who are not depressed? Level of depression would be the independent variable in this study. Level of depression is an individual differences variable and typically is not manipulated by a researcher; rather, the researcher controls this independent variable by systematically selecting individuals who exhibit varying levels of naturally occurring depression. Intelligence, age, aggressiveness, gender, and fraternity or sorority membership are other examples of individual differences variables. Because individuals are selected from preexisting groups, such as those who are depressed, female, aggressive, and so forth, these variables are also sometimes referred to as *natural groups variables*.

The term *independent variable* is used both for individual differences variables when levels of the variable are selected (e.g., level of depression) and for manipulated variables when the levels are implemented by the researcher (e.g., type of booklet describing rape). Nevertheless, as you will see later in Chapter 6, there are important differences between these two kinds of independent variables.

The measures of behavior that are used to assess the effect (if any) of the independent variables are called **dependent variables.** In our example of a study that investigates the effects of alcohol on processing complex information, the researcher might measure the number of errors made by control and experimental participants when playing a difficult video game. Errors, then, would be the dependent variable. Most studies in psychology are done in order to discover the nature of the relationship between particular independent and dependent variables. In the Heath and Davidson (1988) rape study, there were several dependent variables, represented by the women's responses to questions on the rape-perception questionnaire. The results showed that women who read materials emphasizing rape as an uncontrollable event reported higher anxiety when outside on the street or inside their homes, and a greater intention to take precautionary steps, than did women who reviewed the materials describing rape as more controllable. Interestingly, results of additional experiments carried out by Heath and Davidson (1988) indicated that although uncontrollable rape is definitely viewed as anxiety producing, the "intent" to be more cautious reported by their research participants may not always be followed by actual changes in behaviors. Unfortunately, when rape is perceived as an uncontrollable event some women may view attempts to reduce their vulnerability to attacks as futile.

If changes in the dependent variable are to be interpreted unambiguously as a result of the effect of the independent variable, proper control techniques must be used. The story of Clever Hans was used to show how scientists use control to eliminate alternative explanations for a phenomenon (e.g., that Clever Hans was, in fact, clever). Throughout this text, we will be describing specific control procedures used by psychologists when carrying out research (see especially Chapters 6 and 7).

Reporting If we ask someone to *report* to us about events that occurred in our absence, we probably want the report to be unbiased and objective. Otherwise we will be unable to determine exactly what happened. As you might imagine, personal biases and subjective impressions often enter into everyday reports that we receive. Ask anyone to describe an event to you and you are likely to receive not merely details of the event but also personal impressions. You may also find that the details reported to you are not the ones that you would have reported. We often report events in terms of our own interests and attitudes. Obviously, these interests and attitudes do not always coincide with those of others. The next time you take a class examination, poll several classmates on their impressions of the test. Their reports are likely to vary dramatically, depending on such factors as how well prepared they were, what they concentrated on when they studied, and their expectations about what the instructor was going to emphasize on the test.

When scientists report their findings, they seek to separate what they have observed from what they conclude or infer on the basis of these observations. For example, consider the photograph in Figure 1.2. How would you describe to someone what you see there? One way to describe this scene is to say that two people are running along a path with another person running in front of them. You might also describe this scene as three people *racing* each other. If you use this second description, you are reporting an inference drawn from what

FIGURE 1.2 How would you describe this scene?

you have seen and not just reporting what you have observed. The description of three people running would be preferred in a scientific report.

This distinction between description and inference in reporting can be carried to extremes. For example, describing what is shown in Figure 1.2 as running could be considered an inference, the actual observation being that three people are moving their legs up and down and forward in rapid, long strides. Such a literal description also would not be appropriate. The point is that, in scientific reporting, observers must guard against a tendency to draw inferences too quickly. Further, events should be described in sufficient detail without including trivial and unnecessary minutiae. Proper methods for making observations and reporting them will be discussed in Chapter 3.

Scientific reporting seeks to be unbiased and objective. One accepted check on whether a report is unbiased is whether it can be verified by more than one independent observer. A measure of interobserver agreement, for example, is usually found in observational studies (see Chapter 3). Unfortunately, many biases are subtle and not always detected even in scientific reporting. Consider the fact that there is a species of fish in which the eggs are incubated in the mouth of the male parent until they hatch. The first scientist to observe the eggs disappear into their father's mouth could certainly be forgiven for assuming, momentarily, that he was eating them. That's simply what we expect organisms to do with their mouths! But the careful observer waits, watches for unexpected results, and takes nothing for granted.

Concepts We use the term *concepts* to refer to things (both living and inanimate), to events (things in action), and to relationships among things or events, as well as to their characteristics (Marx, 1963). "Dog" is a concept, as is "barking," and so is "obedience." Concepts are the symbols by which we ordinarily communicate. Clear, unambiguous communication of ideas requires that we use concepts that are clearly defined. This means that the concepts we use should be free of unwarranted or surplus meaning. That is, a concept should not convey more meaning than was intended.

In everyday conversation we can often get by without having to worry too much about how we define a concept. Many words, for instance, are commonly used and apparently understood even though neither party to the communication knows *exactly* what the words mean. We are suggesting that people frequently communicate with one another without being fully aware of what they are talking about. This may sound ridiculous but, to illustrate our point, try the following.

Ask a few people whether they believe that intelligence is mostly inherited or mostly acquired. You might try arguing a point of view opposite to theirs just for the fun of it. After having engaged them in a discussion about the roots of intelligence, ask them what they mean by "intelligence." You will probably find that most people have a difficult time defining this concept. Yet people are frequently willing to debate an important point regarding intelligence, and even take a definite stand on the issue, without being able to say exactly what "intelligence" is. When someone does provide a definition, it is unlikely to be exactly the same as that given by another person. That is, "intelligence" means one thing to one person and something else to another. Clearly, in order to attempt to answer the question of whether intelligence is mainly inherited or mainly acquired, we must provide an exact definition that all parties involved can accept. On the other hand, we can talk about it, even argue about it, on an everyday basis without knowing exactly what it is we are talking about!

One way in which a scientist gives meaning to a concept is by defining it operationally. An **operational definition** explains a concept solely in terms of the operations used to produce and measure it. Intelligence, for instance, can be defined operationally by specifying a paper-and-pencil test emphasizing understanding of logical relationships, short-term memory, and familiarity with the meaning of words. Some may not like this operational definition of intelligence, but once a particular test has been identified, there can at least be no argument about what intelligence means *according to this definition*. Operational definitions facilitate communication, at least among those who know how and why they are used.

Although exact meaning is conveyed via operational definitions, this approach to communication has not escaped criticism. One problem has been alluded to already. That is, if we don't like one operational definition of intelligence, there is nothing to prevent us from giving intelligence another operational definition. Does this mean that there are as many kinds of intelligence as there are operational definitions? Each time a new set of questions is added to a paper-and-pencil test of intelligence do we have a new definition of intelligence? The answer, unfortunately, is that we don't really know. To determine whether a different procedure yields a new definition of intelligence, we would have to seek additional evidence. For example, do people who score high on one test also score high on the second test? If they do, the new test may be measuring the same thing as the old one.

Another criticism of using operational definitions is that the definitions are not always meaningful. For example, defining intelligence in terms of how long one can balance a ball on one's nose is an operational definition that most people would not find very meaningful. How do we decide whether a concept has been meaningfully defined? Once again, the solution is to appeal to other forms of evidence. How does performance on a balancing task compare to performance on other tasks that are commonly accepted as measures of intelligence? We must also be willing to apply common sense to the situation. Do people usually consider balancing a ball evidence of intelligence? Scientists are generally aware of the limitations of operational definitions; however, the clarity of communication that derives from this approach is assumed to outweigh the problems it raises.

Instruments You depend on instruments to measure events more than you probably realize. The speedometer in the car, the clock in the bedroom, and the thermometer used to measure body temperature are all instruments that we would find it difficult to do without. And you can appreciate the problems that arise if one of these instruments is inaccurate. *Accuracy* refers to the difference between what an instrument says is true and what is known to be true. A clock that is consistently 5 minutes slow is not very accurate. Inaccurate clocks can make us late, inaccurate speedometers can earn us traffic tickets, and inaccurate thermometers can lead us to believe that we are ill when we are not. The accuracy of an instrument is determined by *calibrating* it, or checking it with another instrument known to be true. Thus, we periodically call the telephone company to check the accuracy of our clocks based on the recorded messages telling us the "time at the tone," which we assume represent the true time. The accuracy of speedometers can be checked using a combination of observations of roadside distance markers and the seconds ticking off on an accurate watch.

Measurements can be made at varying levels of *precision*. A measure of time in tenths of a second is not as precise as one that is in hundredths of a second. One instrument that yields imprecise measures is the gas gauge in most cars. Although reasonably accurate, gas gauges do not give very precise readings. Most of us have wished at one time or another that the gas gauge would permit us to determine whether we had that extra half gallon of gas that would get us to the next service station.

We also need instruments to measure behavior. Wundt used a reaction-time apparatus to measure the time required for cognitive processing. You can be assured that the precision, and even the accuracy, of instruments of this kind have improved significantly in the last 100 years. Today electronic counters provide precise measures of reaction time in milliseconds (thousandths of a second). Many other instruments are employed in contemporary psychology. To perform a psychophysiology experiment (e.g., when assessing a person's arousal level) requires instruments that give accurate measures of such internal states as heart rate and blood pressure. Tests of anxiety sometimes employ instruments to measure galvanic skin response (GSR). Other behavioral instruments are of the paper-and-pencil variety. Questionnaires and tests are popular instruments used by psychologists to measure behavior (see especially Chapter 4). So, too, are the rating scales used by human observers (see Chapter 3). For instance, rating aggression in children on a 7-point scale ranging from not at all aggressive (1) to very aggressive (7) can yield relatively accurate (although perhaps not too precise) measures of aggression. It is the responsibility of the behavioral scientist to use instruments that are as accurate and as precise as possible.

Measurement Psychologists must deal with two types of measurement. The first type, *physical measurement*, involves dimensions for which there is an agreed-upon standard and an instrument for doing the measuring. Length is a dimension that can be scaled with physical measurement, and there are agreed-upon standards for units of length. For instance, 1 meter was at one time defined as 1/10,000,000 of the distance between the North Pole and the equator.

This proved impossible to measure precisely, so the definition of 1 meter was changed to the distance between two points on a platinum-iridium bar kept under controlled conditions near Paris, France. Although this provided accuracy to one part in a million, greater precision was sought. In 1960 the meter was defined as 1,650,763.73 wavelengths of the red-orange radiation of the inert gas krypton-86. But even this proved not precise enough for scientists. (When using the krypton measure to determine the distance from the earth to the moon, scientists found themselves in error by more than 1.5 meters.) The definition of the meter was changed to the length of the path traveled by light in a vacuum in 1/299,792,458 of a second.

In most research in psychology, however, the dimensions to be measured do not involve physical measurement. Rulers do not exist for measuring beauty, aggression, or intelligence. For these dimensions we must use a second type of measurement: *psychological measurement*. Agreement among a certain number of observers provides the basis for psychological measurement. If several independent observers agree that a certain action warrants a rating of 3 on a 7-point rating scale of aggression, we can say that we have a psychological measurement of the aggressiveness of the action. When Albright and Henderson (1995) asked individuals who were depressed and who were not depressed to evaluate other people, the researchers provided participants with an 8-point scale to judge the degree to which various statements about people were true. The rather counterintuitive finding was that depressed individuals provided more accurate ratings of other people than did those who were not depressed.

Of course, many dimensions that can be scaled physically can also be measured using psychological measurement. Observers can be asked to judge which of two lines is longer, and, if they consistently select the longer one, we know the observers can "measure" length. This type of measurement is called psychophysical measurement, and it represents one of the earliest areas of experimental investigation in psychology.

It is important that measurement be both valid and reliable. In general, **validity** refers to the "truthfulness" of a measure. A valid measure of a concept is one that measures what it claims to measure. We discussed this aspect of measurement when we mentioned possible operational definitions of intelligence. Intelligence, it was suggested, could be defined in terms of performance on a task requiring one to balance a ball on one's nose. According to the principle of "operationalism," this is a perfectly permissible operational definition. However, most of us would question whether such a balancing act is really a measure of intelligence. In other words, we would want to know if this is a valid measure of intelligence. Can intelligence actually be measured by how long we can keep a ball balanced on our nose? As we indicated earlier, evidence bearing on the validity of this definition would have to come from other sources. The validity of a measure is supported to the extent that people do as well on it as they do on independent measures that are presumed to measure the same concept. For example, if time spent balancing a ball is a valid measure of intelligence, then a person who does well on the balancing task should also do well

FIGURE 1.3 Balancing a ball on your nose may be an operational definition of intelligence, but is it a valid measure of intelligence?

on such measures as size of vocabulary, reasoning ability, and other accepted measures of intelligence.

The **reliability** of a measurement is indicated by its consistency. Several different kinds of reliability can be distinguished. When we speak of instrument reliability, we are discussing whether an instrument works consistently. The car that sometimes starts and sometimes doesn't when we engage the ignition is not very reliable. Observations made by two or more independent observers are said to be reliable if they show agreement—that is, if the observations are consistent from one observer to another. When several psychologists asked college students to rate the "happiness" of medal winners at the 1992 Summer Olympics in Barcelona Spain, they found that rater agreement was very high (Medvic, Madey, & Gilovich, 1995). They also found, somewhat counterintuitively, that bronze (third place) medal winners were perceived as happier than silver (second place) medal winners, a finding that was explained by a theory of counterfactual thinking. Apparently, people are happier just making it (to the medal stand) than they are just missing it (i.e., missing a gold medal).

In the context of psychological testing, researchers need to be concerned with test reliability. Are the results obtained on a test consistent from one administration of the test to another? When test scores are not reliable, we don't know what a particular score means. Suppose that in preparing to apply to graduate school in psychology you take the Graduate Record Examination (GRE), a test that is required of applicants to many graduate programs. Scores on this test are then sent to schools to which you are applying. Only if the GRE is a reliable test can a school trust that your score represents how you consistently perform on tests like this one. If a test is not reliable, then we have no way of knowing whether a particular score is high, low, or somewhere in the middle of scores you might possibly obtain on this test.

In subsequent chapters of this book, we will discuss various kinds of validity and reliability and introduce you to various methods for measuring them.

Hypotheses A **hypothesis** is a tentative explanation for something. It frequently attempts to answer the questions "How?" and "Why?" At one level, a hypothesis may simply suggest how particular variables are related. For example, consider the phenomenon of "depressive realism," which we have alluded to previously. This refers to the idea that in many situations depressed individuals make more accurate (or realistic) judgments than do individuals who are not depressed. Albright and Henderson (1995) wondered whether depressive realism would extend to depressives' judgments about people. They hypothesized that depressed individuals would judge other people more accurately than would those who were not depressed. That is, they hypothesized that level of depression and accuracy of judgments were *related*. You may remember that when these researchers asked participants to rate the degree to which various statements about people were true, the results supported their hypothesis: Depressed individuals gave more accurate ratings of other people than did those who were not depressed.

At a more theoretical level, a hypothesis may offer a reason (the "why") for the way that particular variables are related. For example, Albright and Henderson (1995) suggested that many of us avoid symptoms of depression by perceiving other people less favorably than is actually true. (This way we always look good!) Such a strategy may help us to feel good about ourselves. These researchers further hypothesized that depressed individuals, in contrast, may lose this ability to view people unfavorably. That is, depressed individuals make more accurate (realistic) judgments *because* they have lost the ability to view others unfavorably. Moreover, consistent with their reasoning, individuals who were not depressed (unlike depressed individuals) rated others less favorably than was objectively true.

Nearly everyone has proposed hypotheses to explain some human behavior at one time or another. Why do people commit apparently senseless acts of violence? What causes people to start smoking cigarettes? Why are some students academically more successful than others? One characteristic often distinguishes the hypotheses proposed by the nonscientist from those offered by the scientist: testability. As attractive as it might be, if a hypothesis cannot be tested, it is of no immediate use to science (Marx, 1963).

Hypotheses are not testable *if the concepts to which they refer are not adequately defined*. To say that a would-be assassin shot a U.S. president or other prominent figure because he was mentally disturbed is not a testable hypothesis unless a definition of mentally disturbed can be agreed upon. Unfortunately, psychologists and psychiatrists cannot always agree on what terms such as "mentally disturbed" mean. Often this occurs because an accepted operational definition is not available for these concepts. You may have learned in a psychology class that many of Freud's hypotheses are not testable. One reason has to do with the lack of operational definition for such concepts as *id*. As one prominent researcher has commented, "a criterion of whether or not a so-called empirical concept is a scientific concept is whether or not it has been operationally defined" (Underwood, 1957, p. 52). Therefore, in addition to facilitating clarity in communication, operational definitions offer a means of evaluating whether our hypotheses contain scientifically acceptable concepts.

Hypotheses are also untestable if they are *circular*, in which case the event itself becomes an explanation for the event. One scientist has pointed out that using a definition of an event as an explanation for the event—that is, a *circular hypothesis*—is something you can "catch the late-night talk show hosts doing" all the time (Kimble, 1989). Consider these examples:

— Your eight-year-old son is distractable in school and having trouble reading because he has an attention deficiency disorder.
— The stock market crash of October 19, 1987, was caused by widespread economic panic. (Kimble, 1989, p. 495)

Do you see that by saying a boy is "distractable . . . *because* he has an attention deficiency disorder" doesn't explain anything? An attention deficiency disorder is defined by the inability to pay attention. Thus, such a statement simply says that he doesn't pay attention because he doesn't pay attention! Or that by saying a "stock market crash . . . was *caused* by widespread economic panic" simply suggests that economic panic (as defined by the stock market crash) is due to economic panic?

This scientist goes on to say that "If [he] could make just one change in what the general public (and some psychologists) understand about psychology, it would be to give them an immunity to such misuses of definitions" (Kimble, 1989, p. 495).

A hypothesis also may be untestable if it *appeals to ideas or forces that are not recognized by science*. As we have shown, science deals with the observable, the demonstrable, the empirical. To suggest that people who commit horrendous acts of violence are under orders from the Devil is not testable because it invokes a principle (the Devil) that is not in the province of science. Such hypotheses might be of value to philosophers or theologians but not to the scientist.

Hypotheses in scientific research are often derived from a theory. We'll discuss the nature of scientific theories, their function, and how they are constructed and tested later in this chapter. But now, let us complete this description of a scientific approach by discussing the attitude you must have as a scientist.

Attitude More than anything else, scientists are skeptical. Not only do they want to "see it before believing it," but they are likely to want to see it again, and again, perhaps under conditions of their own choosing. Behavioral scientists come to this skepticism by recognizing two important facts. First, behavior is complex, and often many factors interact to give rise to a psychological phenomenon. Discovering these factors is often a difficult task. The explanations proposed are sometimes premature, not enough factors having been considered or the existence of one or more factors having gone unnoticed. Second, the behavioral scientist recognizes that science is a human endeavor. People make mistakes. Human inference, as we have suggested, is not always to be trusted. Therefore scientists are often skeptical about "new discoveries" and extraordinary claims.

The skepticism of scientists produces a cautiousness that is often lacking in those without scientific training. Too many people are apparently all too ready to accept explanations that are based on insufficient or inadequate evidence. This is illustrated by the widespread belief in the occult. Rather than approach cautiously the claims of those who promote belief in the paranormal, many people are uncritical in their acceptance. According to public opinion surveys, a large majority of Americans believe in ESP (extrasensory perception), and many people apparently are convinced that beings from outer space have visited earth. About 2 in 5 Americans give some credibility to astrology reports, and as many as 12 million adults report changing their behavior after reading astrology reports (Miller, 1986). Such beliefs are held despite minimal and often contradictory evidence of the validity of horoscopes. This human tendency to ignore certain kinds of evidence is not new. When scientists successfully demonstrated the means by which Clever Hans was so clever, many people continued to believe in his superior reasoning ability (Watson, 1914/1967).

What is responsible for this tendency to propose explanations based on the occult and to resist obvious evidence to the contrary? Singer and Benassi (1981) suggest that an uncritical attitude toward such events has several sources. First, there are many distortions in the media. The public is constantly exposed to television shows, newspaper accounts, and other reports of events presumably caused by supernatural forces. These reports are often presented with little critical evaluation, and scientific evidence for alternative explanations is frequently ignored. The sheer pervasiveness of such reports may lend credibility to them.

Another reason for the widespread acceptance of the occult may be deficiencies in human reasoning. As we emphasized earlier in this chapter, everyday inferences are susceptible to many biases, including the tendency to seek only confirmatory evidence, to jump to conclusions, and to perceive causality in events when none is actually present (see, for example, Zechmeister & Johnson, 1992). Finally, Singer and Benassi find fault with science education. Too often, they suggest, science is taught as a set of facts rather than as a way of approaching knowledge critically and systematically. Thus, people are often impatient with the scientific process and even confused when scientists appear to change their minds or when they attempt to clarify earlier findings. Singer and Benassi also find that the general public exhibits woefully little general scientific knowledge. Many people believe that islands float on the ocean surface, for instance, and that the moon is fixed in the sky but only visible at night. Belief in astrology reports and "lucky numbers," for example, is stronger among people with little formal education than it is among college-educated individuals (Miller, 1986).

Scientists do not, of course, automatically assume that unconventional interpretations of unexplained phenomena could not possibly be true. They simply insist on being allowed to test all claims and to reject those that are inherently untestable. Scientific skepticism is a gullible public's defense against charlatans and others who would sell them ineffective medicines and cures, impossible schemes to get rich, and supernatural explanations for natural phenomena.

Goals of the Scientific Method

- The scientific method is intended to meet three goals: description, prediction, and understanding.
- Psychologists seek to describe events and relationships between variables; most often, researchers use the nomothetic approach and quantitative analysis.
- Correlational relationships allow psychologists to predict behavior or events, but do not allow psychologists to infer what causes these relationships.
- Psychologists understand the cause of a phenomenon when the three conditions for causal inference are met: covariation, time-order relationship, and elimination of plausible alternative causes.
- The experimental method, in which researchers manipulate independent variables to determine their effect on dependent variables, establishes time-order and allows a clearer determination of covariation.
- Plausible alternative causes for a relationship are eliminated if there are no confoundings in a study; a study free of confoundings has internal validity.
- External validity refers to the extent to which a study's findings may be used to describe different populations, settings, and conditions.

Description Description refers to the procedures by which events and their relationships are defined, classified, cataloged, or categorized. Clinical research, for instance, has provided practitioners with many different sets of criteria for classifying mental disorders. Many of these are found in the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., 1994), also known as *DSM-IV*. Consider, as one example, the criteria used to define the disorder labeled *Dissociative Fugue* (formerly Psychogenic Fugue).

Diagnostic Criteria for Dissociative Fugue

A The predominant disturbance is sudden, unexpected travel away from home or one's customary place of daily activities, with inability to recall one's past.
B Confusion about personal identity or assumption of a new identity (partial or complete).

C The disturbance does not occur exclusively during the course of Dissociative Identity Disorder and is not due to the direct physiological effects of a substance (e.g., a drug of abuse, medication) or a general medical condition (e.g., temporal lobe epilepsy).

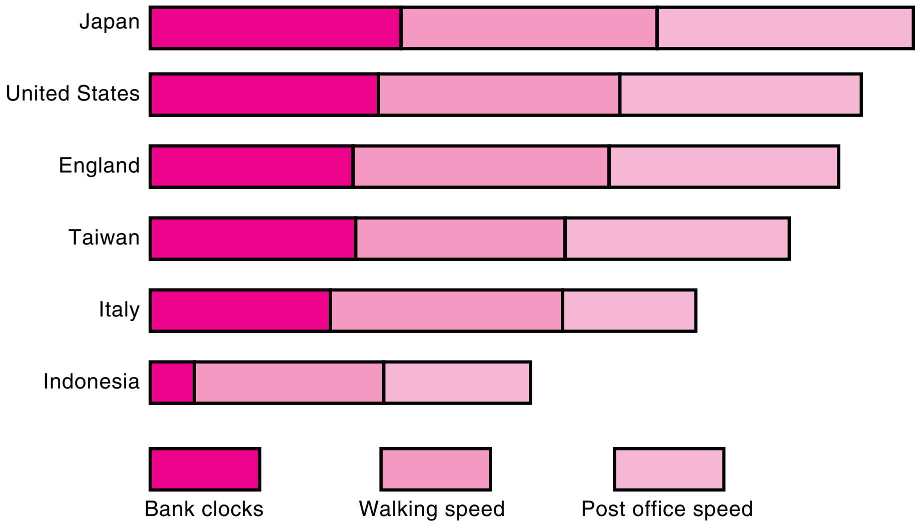
D The symptoms cause clinically significant stress or impairment in social, occupational, or other important areas of functioning. (*DSM-IV*, 1994, p. 481)

The diagnostic criteria used to define dissociative fugue provide an operational definition for this disorder. Like many other unusual mental disorders, dissociative fugues are relatively rare; thus, we typically learn about these kinds of disorders based on individual descriptions of people exhibiting them. These descriptions are called "case studies." Clinical research frequently makes use of case studies, a procedure that will be discussed in detail in Chapter 9. Research also seeks to provide clinicians with descriptions of the prevalence of a mental disorder and their relationship to variables including, but not restricted to, gender and age. According to the *DSM-IV* (1994), for instance, dissociative fugue is seen primarily in adults, and although it is relatively rare, it is more frequent "during times of extremely stressful events such as wartime or natural disaster" (p. 482).

Another example of description in psychological research is the work of Levine (1990) who described the "pace of life" in various cultures and countries of the world. Measures of a country's tempo were made by noting the accuracy of outdoor bank clocks in a country's cities, by timing the walking speed of pedestrians over a distance of 100 feet, and by measuring the speed with which postal clerks processed a standard request for stamps. The investigator frequently enlisted the help of students to make observations while they traveled during summer vacations or semester breaks. The results of this study are shown in Figure 1.4. The citizens of Japan exhibited, overall, the fastest pace of life; the citizens of Indonesia were the slowest. U.S. citizens were second overall.

Psychology (like science in general) develops descriptions of phenomena using the *nomothetic approach*. The objective of the nomothetic approach is to establish broad generalizations and "universal laws" that apply to a wide population of organisms. As a consequence, psychological research frequently consists of studies involving large numbers of participants with the purpose of determining the "average," or typical, performance of a group. This average may or may not represent the performance of any one individual in the group. Not all citizens of Japan or the United States, for example, are on the fast track. In fact, Levine (1990) and his colleagues found wide differences in the pace of life among various cities within a country. Inhabitants of large cities walked faster than did those of medium-sized cities in the various countries that were visited. In the United States, differences in a city's tempo were found depending on the region of the country. Cities in the Northeast (e.g., Boston, New York) had a faster tempo than did cities on the West Coast (e.g., Sacramento, Los Angeles). Of course, there will be individual variations within cities as well. Not all citizens of Los Angeles are going to be slow-paced, nor are all New Yorkers going to be fast-paced. Nevertheless, the Japanese move *in general* at a faster

FIGURE 1.4 Measures of accuracy of a country's bank clocks, pedestrian walking speed, and the speed of postal clerks performing a routine task served to describe the pace of life in a country. In the graph a longer bar represents greater accuracy of clocks or greater speed of walking and performing a task (from Levine, 1990).



pace than do Indonesians, and Americans on the West Coast exhibit, *on the average*, a slower pace of life than do residents of the Northeast.

The nomothetic approach does not deny that there are important differences *among* individuals; it simply seeks to identify the similarities that exist among these differences. For example, a person's individuality is not threatened by our knowledge that that person's heart, like the hearts of other human beings, is located in the upper left chest cavity. Similarly, we do not deny a person's individuality when we state that that person's behavior is influenced by patterns of reinforcement. Researchers merely seek to describe what organisms are like in general on the basis of the average performance of a group of different organisms.

Although the nomothetic approach predominates in psychological research, there is an alternative. Some psychologists, notably Allport (1961), argue that the nomothetic approach is inadequate—that the individual cannot be represented by an average value. Allport argues that the individual both is unique and exhibits behavior that conforms to general laws, or principles. He maintains that study of the individual, called *idiographic research*, is important. A major form of idiographic research is the case study method, which we will describe in detail in Chapter 9.

Researchers decide whether to make generalizations about groups of individuals or to look for the lawfulness found in one individual's behavior. This decision is largely dictated by the nature of the question being asked. And although many researchers do mainly one or the other kind of research, others may do both. A clinical psychologist, for instance, may decide to pursue mainly idiographic investigations of a few clients in therapy but consider nomothetic issues when doing research with groups of college students. Another decision that the researcher must make is whether to do quantitative or qualitative research. *Quantitative research* refers to studies whose findings are mainly the product of statistical summary and analysis. *Qualitative research* produces research findings that are not arrived at by statistical summary or analysis and lack quantification altogether (see, for example, Strauss & Corbin, 1990).

The data of qualitative research are most commonly obtained from interviews and observations and can be used to describe individuals, groups, and social movements (Strauss & Corbin, 1990). Qualitative research is often about "naturally occurring, ordinary events in natural settings" (Miles & Huberman, 1994, p. 10). Powell-Cope (1995), for instance, used qualitative methods to "describe the experiences of gay couples when at least one was diagnosed with symptomatic HIV infection or AIDS" (p. 36). Van Meter, Yokoi, and Pressley (1994) used a qualitative research approach to develop a theory of student note-taking. The researchers met with university undergraduates in small discussion groups, conducted both directed and undirected interviews, then developed and administered formal questionnaires. According to the authors, this qualitative approach to research on student note-taking yielded "many insights into note-taking dynamics that have not been identified in previous research" (p. 323). Other examples of qualitative research are found in Chapter 3 when we discuss narrative records of observed behavior; case studies described in Chapter 9 also are a form of qualitative research. Just as psychological research is more frequently nomothetic than idiographic, it is also more typically quantitative than qualitative. And although both kinds of research can be usefully employed to describe behavior, our emphasis in this book is mainly on quantitative research.

Prediction Description of events and their relationships often provides a basis for *prediction,* the second goal of the scientific method. There are important questions in psychology that call for predictions. For example: Does the early loss of a parent make a child especially vulnerable to depression? Are children who are overly aggressive likely to have emotional problems as adults? Do stressful life events lead to increased physical illness? Research findings suggest an affirmative answer to all of these questions. This information not only adds valuable knowledge to the discipline of psychology but is also helpful in both the treatment and the prevention of emotional disorders.

An important occupation of many psychologists is the prediction of later performance (e.g., on the job, in school, or in specific vocations) on the basis of earlier performance on various standardized tests. For instance, we mentioned earlier the Graduate Record Examination or GRE. Scores on this test, as well as undergraduate grade-point average (GPA), can be used to predict how well a student will do in graduate school. Yet Sternberg and Williams (1997) suggest that there are limitations on this prediction. They found that while GRE scores predicted fairly well the first-year grades of graduate students at their institution, the GRE was not predictive of other, important performance criteria such as advisors' ratings of a student's creativity, ability to teach, and ability to do research. Not surprisingly, these researchers have sparked a debate by questioning the validity of the GRE, which is so widely regarded as a predictor of students' later professional development (see, for example, "Comment" section of *American Psychologist,* 1998, *53,* 566–577).

Interestingly, research has shown that faculty recommendations, which are usually required for graduate school admittance, are also a rather poor predictor of whether a student will successfully complete a doctorate degree (Willingham, 1974). On the other hand, there is research showing that the amount of undergraduate research activity and ratings by peers of a student's commitment to psychology are better predictors of later success in psychology (as measured, for example, by number of scientific publications) than are the more traditional measures, such as GRE scores and GPA (Hirschberg & Itkin, 1978).

When scores on one variable can be used to predict scores on a second variable, we say that the two variables are correlated. A **correlation** exists when two different measures of the same people, events, or things vary together—that is, when particular scores on one variable tend to be associated with particular scores on another variable. When this occurs, the scores are said to "covary." Consider a measure with which you likely have had some experience, namely teacher/course evaluations in classes that you have taken. College students are commonly asked to evaluate their instructors and the course material toward the end of a course. By the time a course is over, you probably have formed many impressions of a teacher (e.g., whether the instructor is supportive, enthusiastic, likeable). After all you have just spent as many as 12 or 14 weeks (perhaps more than 30 hours) in this instructor's classroom. Ambady and Rosenthal (1993) asked how well teacher evaluations by students *not* enrolled in the class would correlate with end-of-the-semester evaluations made by students in the class. They showed video clips (without sound) of teachers to a group of female undergraduates. But, and here is the interesting part, they showed the video clips for only 30 seconds, 10 seconds, or just 6 seconds (across several studies). The researchers found that teacher evaluations based on these "thin slices of nonverbal behavior" correlated well with end-of-the-semester teacher evaluations made by students who were enrolled in the class. These results indicate that people (in this case teachers) reveal much about themselves when their nonverbal behavior is seen only briefly, and also that we (as observers) can make relatively accurate judgments of affective behavior quite quickly. Ambady and Rosenthal's findings, of course, do not mean that all the information in teaching evaluations can be captured by this method as they focused only on judgments of affective behavior (e.g., likeableness).

We can cite one more example of correlational evidence from a study we introduced earlier. When studying the "pace of life," Levine (1990) found a positive correlation between a city's pace of life as measured by people's walking speed and its deaths from heart disease. This indicates that the faster the pace in a city, the more likely its inhabitants are to die from heart disease. Death rates from heart disease, in other words, can be predicted by simply measuring how fast a city's inhabitants typically walk 100 feet!

We are concerned with issues of correlation at various points in this text and discuss the more general topic of correlational research in Chapter 4. Correlational research, which seeks to describe predictive relationships among variables, is a major area of psychological research.

It is important to point out that successful prediction doesn't always depend on knowing why a relationship exists between two variables. Consider the report that the Chinese rely on observing animal behavior to help them predict earthquakes. Certain animals apparently behave in an unusual manner just before an earthquake. The dog that barks and runs in circles and the snake seen fleeing its hole, therefore, may be reliable predictors of earthquakes. If so, they could be used to warn people of forthcoming disasters. We might even imagine that in areas where earthquakes are likely, residents would be asked to keep certain animals under observation (as miners once kept canaries) to warn them of conditions of which they are as yet unaware. This would not require that we understand why certain animals behave strangely before an earthquake, or even why earthquakes occur.

You may remember that Levine (1990) showed that measures of the pace of a city can be used to predict death rates from heart disease. However, we can only speculate about why these measures are related. One possible explanation for this correlation suggested by the researchers is that people living in time-urgent environments engage in unhealthy behaviors, for example, cigarette smoking and poor eating habits, which increase their risk of heart disease (Levine, 1990). To account for the relationship between *non*student ratings and student teacher evaluations, Ambady and Rosenthal (1993) suggested that people are "attuned" to quickly picking up information about a person's affect because this information is important (adaptive) in real-life decision making. Without additional information, however, this idea must remain speculation.

Understanding Although they are important goals in themselves, description and prediction are only the first steps in understanding a phenomenon. Understanding is the third and most important goal of the scientific method. It is achieved when the cause or causes of a phenomenon are identified. The scientist sets three important conditions for making a **causal inference** and, hence, of understanding: *covariation of events; a time-order relationship; and, the elimination of plausible alternative causes.*

If one event is the cause of another, the two events must vary together—that is, when one changes, the other must also change. This is the principle of covariation that we noted was the basis of the goal of prediction. Further, the presumed cause must occur before the presumed effect. This is the second condition, a time-order relationship. Finally, causal explanations are accepted only when other possible causes of the effect have been ruled out—when plausible alternative causes have been eliminated.

Unfortunately, people have a tendency to conclude that all three conditions for a causal inference have been met when really only the first condition is satisfied. For example, research has suggested that parents who are stern disciplinarians and who use physical punishment are more likely to have aggressive children than are parents who are less stern and use other forms of discipline. Parental discipline and children's aggressiveness obviously covary. Moreover, the fact that parents are typically assumed to influence how their children behave might lead us to think that the time-order condition has been met—parents use physical discipline and children's aggressiveness results. More recent research has shown, however, that infants vary in how active and aggressive they are and that the infant's behavior has a strong influence on the *parents'* responses in trying to exercise control. In other words, some children may be naturally aggressive and require stern discipline rather than stern discipline producing aggressive children. Therefore, the direction of the causal relationship may be opposite to what we thought at first.

It is important to recognize, however, that the causes of events cannot be identified unless covariation has been demonstrated. The first objective of the scientific method, description, can be met by describing events under a single set of circumstances. The goal of understanding, however, requires more than this. For example, suppose that a teacher wished to demonstrate that so-called "active learning strategies" help students learn. She could teach students using this approach and then describe the performance of the students who received instruction in this particular way. But at this point what would she know? How do we know that another group of students taught using an approach that did not use active learning strategies might not learn the same amount? Before the teacher could claim that the performance she observed was caused by the use of active learning strategies, she would have to compare this method with some other reasonable approach that did not employ active learning strategies. She would look for a difference in learning between the group using active learning strategies and a group not using this method. Such a finding would show that teaching strategy and performance covary. As you will see in Chapter 6, this is one reason why manipulation is such an important method of control for scientists. By manipulating independent variables, such as type of learning strategy, the investigator can determine more clearly whether any subsequent covariation occurs in the dependent variable. If it does, a bonus results from using manipulation: The time-order condition has also been met, because the researcher changes the independent variable (e.g., teaching method) and *subsequently* measures the changes, if any, in the dependent variable (e.g., a measure of student learning).

By far the most challenging condition to be met in making a causal inference is eliminating other plausible alternative causes. For example, consider a study in which the effect of two different teaching approaches (active and passive) is to be assessed. You can see that one way not to assign students to teaching conditions in this study would be to have all men in one group and all women in the other. If this were done, the independent variable of teaching method would be "confounded" with the independent variable of gender. **Confounding** occurs when two potentially effective variables are allowed to covary simultaneously. When research is confounded, it is impossible to determine what variable is responsible for any obtained difference in performance. If males were assigned to one condition of our hypothetical teaching experiment and females to another, then we would not know whether any difference we observed was due to gender of the participant or to the independent variable of type of teaching. When no confoundings are present, a research study is said to have **internal validity.** Much of the discussion in this book will focus on developing skills in performing internally valid research by learning to identify and eliminate confoundings.

The internal validity of a study, which relates to our ability to make causal inferences based on it, must be distinguished from the study's **external validity.** External validity involves the extent to which the research results can be generalized to different populations, settings, and conditions. A research finding may be internally valid and have little external validity. Although free of confoundings, the study may not be able to be generalized. Questions of external validity in science often arise, for instance, when results obtained with animal subjects are claimed to apply to human participants as well. Researchers frequently depend on experiments with animals to obtain information about human psychopathology and other illnesses. Animal models have been used to study drug addiction, minimal brain dysfunction, and various forms of mild to severe psychopathology (Maser & Seligman, 1977), as well as psychosomatic factors in various human illnesses (see Hofer & Myers, 1996). A frequent focus of animal research has been the effect of stress on the development of mental and physical illness. And while these studies have yielded valuable information about the effects of stress in animals, it is not clear whether these findings can be easily generalized to situations involving human responses to stress (e.g., Weiner, 1996). Nevertheless, as Hofer and Myers (1996) emphasize, the goal of animal research is the "enrichment of clinical and human research," and while "in themselves animal studies can never settle questions about human nature, . . . they can allow clinicians and those who study normal humans to ask more highly focused research questions that can be tested in the limited studies that are possible with our own species" (p. 522).

Because laboratory research, in general, is often conducted under more controlled conditions than are found in natural settings, an important task of the scientist is to determine whether laboratory findings generalize to the "real world." This is often the goal of what is called *field research*. By performing in the "field" (outside the laboratory) an experiment that is similar to an experiment that has been carried out in the laboratory, a researcher can provide evidence bearing on the external validity of laboratory results. Research conducted outside the laboratory often requires methods and procedures that are specifically designed for this less well controlled environment. Later in this book we will introduce some of the methods used by psychologists to conduct field research (see especially Chapters 3 and 10).

Scientific Theory Construction and Testing

- Theories, proposed explanations for the causes of behavior, vary in their scope and their level of explanation.
- A scientific theory is a logically organized set of propositions that defines events, describes relationships among events, and explains the occurrence of events.
- Successful scientific theories organize empirical knowledge, guide research by offering testable hypotheses, and survive rigorous testing.
- Theories frequently posit intervening variables to explain observed behavior.
- Researchers evaluate theories by judging the theory's internal consistency, observing whether hypothesized outcomes occur when the theory is tested, and noting whether the theory makes precise predictions based on parsimonious explanations.

Theories are "ideas" about how nature works. Psychologists propose ideas about why behavior occurs the way it does (e.g., what causes schizophrenia), about the nature of cognition (e.g., how people solve problems), and so on. A psychological theory can be developed on different levels: For example, it can be developed on a physiological or on a symbolic level (see Anderson, 1990; Simon, 1992). A theory of schizophrenia may, in other words, propose biological causes (e.g., specific genetic carriers), psychological causes (e.g., patterns of emotional conflict, stress), or both. The propositions contained in theories may be expressed as verbal statements, as mathematical equations, or even as computer programs.

An important dimension on which theories differ is scope. The *scope* of a theory refers to the range of phenomena that it seeks to explain. A theory of "flashbulb memory," for instance, attempts to explain why the personal circumstances surrounding particularly surprising and emotional events, such as the explosion of the space shuttle *Challenger*, are remembered better than are details associated with everyday events (see, for example, Brown & Kulik, 1977; McCloskey, Wible, & Cohen, 1988). Many older adults, for instance, can apparently remember precisely what they were doing when they heard about the assassination of President Kennedy. The scope of a theory of flashbulb memory is relatively restricted, proposing as it does an explanation of the nature and cause of an interesting but very specific memory phenomenon. A theory of human love, such as that proposed by Sternberg (1986), is an example of a psychological theory of much greater scope. Sternberg shows how the amount and kind of love one experiences is a function of three critical behavioral components: intimacy, passion, and decision/commitment. Even broader in scope is Anderson's (1990; 1993; Anderson & Milson, 1989) theory of human cognition, which offers an account of how cognition works in general, including ideas about learning, memory, problem solving, and so on. Clearly, the scope of a theory can be quite large. In general, the greater the scope of a theory the more complex it is likely to be. Complexity may be a necessary characteristic of psychological theories

FIGURE 1.5 A theory of "flashbulb memory" seeks to explain why personal circumstances surrounding vivid, emotional events, such as the explosion of the space shuttle *Challenger*, are seemingly remembered so much better than are details of everyday events.

given the nature and range of phenomena psychologists try to understand. Complexity also can be a serious obstacle, however, to testing a theory. Most theories in contemporary psychology tend to be relatively modest in scope, attempting to account only for a limited range of phenomena.

The source of a scientific theory is a mixture of intuition, personal observation, and discovered knowledge (known facts and ideas). The famous philosopher of science, Karl Popper (1976, pp. 268–269), suggested that truly creative theories spring from a combination of "intense interest in a problem (and thus a readiness to try again and again)" and "critical imagination." *Critical imagination* is the ability to think critically, but there is more. It includes a readiness to challenge accepted ideas and an "imaginative freedom that allows us to see so far unsuspected sources of error" in previous thinking about a problem. Critical imagination, in other words, means traveling beyond what others have said are the boundaries (limits) of thinking about a problem. Assuming we have that important burning interest in a problem, one way we might approach constructing a scientific theory is by critically examining what is known, looking for flaws or unseen sources of error in that knowledge.

Whatever the nature and scope of a theory, whether it be expressed mathematically or verbally or developed at a physiological level or at some higher-order level, the theory includes certain assumptions and concepts that must be explained in order for it to be understood and tested. A theory of flashbulb memory, for instance, needs to state exactly what a flashbulb memory *is*, showing, for example, how flashbulb memory differs from other, more typical memory. A complete theory must also explain why in some cases a person's so-called flashbulb memory is clearly wrong, even though the individual expresses high confidence in the (inaccurate) memory (see Neisser & Harsch, 1992). To be complete, therefore, a theory must include definitions of various events or concepts (e.g., emotional events, test accuracy), information about relationships between these events (e.g., the relationship between degree of emotional involvement and amount remembered), and so forth. Thus, we can offer the following formal definition of a scientific **theory:** *a logically organized set of propositions (claims, statements, assertions) that serves to define events (concepts), describe relationships among these events, and explain the occurrence of these events*.

The major functions of a theory are to *guide* research and to *organize* empirical knowledge (Marx, 1963). In the early 1960s, Rotter (1966) developed a theory of internal versus external locus of control. In this theory, Rotter differentiates individuals who perceive a contingency between their behavior and what happens to them (internal locus of control) and those who perceive that their behavior has little consequence for what happens to them (external locus of control). Rotter developed a questionnaire to define this concept on the basis of the idea that individuals differ in locus of control because of the way they have been rewarded during their development. The theory suggests a relationship between perceived locus of control and anxiety, with greater anxiety associated with greater perceived external locus of control. The theory has guided researchers for many years and has served to organize a body of empirical literature regarding self-efficacy, or the feeling of being able to cope with the environment (see, for example, Myers, 1999). The success of a psychological theory such as the locus-of-control theory can be measured by the degree to which it achieves the two important goals of guiding research and organizing empirical findings.

A scientific theory guides research by suggesting testable hypotheses. You may remember that hypotheses, like theories, are explanations for behavior; however, a hypothesis typically is simpler and more tentative than is a scientific theory (Marx, 1963). We described earlier in this chapter several factors that affect the testability of hypotheses, including the important criterion of operational specificity of concepts. It is important that hypotheses derived from a theory attempt to meet these criteria.

Theories frequently require that we postulate intervening processes to account for observed behavior (Underwood, 1975). An intervening process is one that "mediates" between certain antecedent conditions and behavior. We have been discussing one such intervening process, memory. We can safely say that we have never "seen" memory. It is something that we infer based on observations of our own behavior and that of others when asked about a previous experience. We assume that there is some record of experiences somewhere inside us which we refer to when queried about a previous experience. The concept "memory" is proposed as an intervening process to explain the behavior we have observed. Mediating processes such as memory are called *intervening variables*. To be useful to the scientist, it is important that these concepts have a clear relationship to behavior (see Kimble, 1989). That is, although we might infer many kinds of unseen processes to help explain behavior, unless we tie down intervening processes to observed relationships between specific antecedents (independent variables) and behavior (dependent variables), our theories are scientifically weak. For example, a theory that proposes a "free-floating" process to explain behavior cannot be shown to be wrong. Because free-floating processes (e.g., "the little person inside me") are not tied to specific empirical relationships, they can be called into play in any and every situation to "explain" behavior.

How scientific theories should be evaluated and tested is one of the most complex and difficult issues in psychology and philosophy (e.g., Meehl, 1978, 1990a, 1990b; Popper, 1959). At least one scientist, however, suggests a rather straightforward procedure. Kimble (1989, p. 498) says quite simply, "The best theory is the one that survives the fires of logical and empirical testing." Although somewhat simplistic, this is a good starting point. A theory can be evaluated first on the basis of its organization, its logical consistency. Are its propositions arranged in a meaningful way? Is the theory free of contradictions? Can specific deductions about expected behavior be made from it easily? Theories are tested logically by exposing their internal structure to the critical eye of members of the scientific community. A particular theory's propositions, assumptions, and definitions are frequently the topic of debate in scientific journals. Ideas about the exact definition of flashbulb memories, for example, have been debated at length in the psychology literature (e.g., Cohen, McCloskey, & Wible, 1990; Pillemer, 1990; Winograd & Neisser, 1992).

Kimble (1989) suggests that a theory is strengthened or weakened according to the outcomes of empirical tests of hypotheses derived from it. Successful tests of a hypothesis serve to increase the acceptability of a theory; unsuccessful tests serve to decrease the theory's acceptability. The best theory, in this view, is the one that passes these tests successfully. But there are serious obstacles to testing hypotheses and, as a consequence, confirming or disconfirming scientific theories. For example, a theory, especially a complex one, may produce many specific testable hypotheses. Therefore, a theory is not likely to fall on the basis of a single test (e.g., Lakatos, 1978). Moreover, theories may include propositions and concepts that have not been adequately defined or suggest intervening processes that are related to behavior and to each other in complex and even mysterious ways. Such theories may have a long life, but their value to science is questionable (Meehl, 1978).

When constructing and evaluating a theory, scientists place a premium on parsimony (Marx, 1963). The *rule of parsimony* is followed when the simplest of alternative explanations is accepted. When choosing among theoretical propositions, scientists tend to favor the simplest of them. This rule is nicely illustrated by the comments made by Albert Einstein in response to a young psychology student's question about the cognitive "process that occurred when

[he] had [his] most productive thought" (Fromm, 1998, p. 1195). In 1932, Erika Fromm, who was then a student of Max Wertheimer, the founder of Gestalt psychology, wrote letters to 100 leading scientists and philosophers as part of a psychological study of productive thinking. Among the responses she received was one from Einstein. He described the nature of the thinking that led to his theory of relativity, a process which he then characterized in the following way:

> It was always the search for a logically simple meaning of empirically established relationships, propelled by the conviction that there existed a simple logical meaning. (Fromm, 1998)

Precision of prediction is another criterion by which a theory can be evaluated. Theories that make precise predictions about behavior are preferred to those that make only general predictions (Meehl, 1990a). For instance, a theory of flashbulb memory that predicts the precise nature and duration of a "forgetting function" for such a memory is clearly a better theory than one that simply states that these memories will be remembered "longer" than other memories. Stated another way, some tests are easier for theories to pass than are others. A good scientific theory is one that is able to pass the most rigorous tests. *Rigorous testing* includes more tests that seek to falsify a theory's propositions than ones that seek to confirm them (Cook & Campbell, 1979). While confirming a particular theory's propositions provides support for the specific theory that is being tested, confirmation logically does not rule out other, alternative theories of the same phenomenon.

Although theories can be difficult and challenging to work with, the process of constructing and evaluating scientific theories is at the core of the scientific enterprise and is absolutely essential for the continuation of the science of psychology.

Scientific Integrity

- Psychologists are obligated to conduct research of the highest integrity.
- The American Psychological Association (APA) has adopted a code of ethics to guide the professional behavior of psychologists.

Science is a search for truth. Fraud, lies, and misrepresentations should play no part in a scientific investigation. But science is also a human endeavor, and frequently much more is at stake than truth. Both scientists and the institutions that hire them compete for rewards in a game with jobs, money, and reputations on the line. The number of scientific publications authored by a university faculty member, for instance, is usually a major factor influencing decisions regarding professional advancement through promotion and tenure. Under these circumstances, there are unfortunate, but seemingly inevitable, cases of scientific misconduct.

A variety of activities constitute violations of scientific integrity. They include data fabrication, plagiarism, selective reporting of research findings, failure to acknowledge individuals who made significant contributions to the research, misuse of research funds, and unethical treatment of humans or animals (see Adler, 1991). Some transgressions are easier to detect than others. Out-and-out fabrication of data, for instance, can be revealed when, in the normal course of science, results are not able to be reproduced by independent researchers, or when logical inconsistencies appear in published reports. However, more subtle transgressions, such as reporting only data that meet expectations or misleading reports of results, are difficult to detect. The dividing line between intentional misconduct and simply bad science is also not always clear.

To educate researchers about the proper conduct of science, and to help guide them around the many ethical pitfalls that are present, most scientific organizations have adopted formal codes of ethics. In Chapter 2 we will introduce you to the APA ethical principles governing research with humans and animals. As you will see, ethical dilemmas often arise. Consider the research by Heath and Davidson (1988) mentioned earlier in this chapter. You may remember that they asked groups of university women to help prepare a new rape-prevention pamphlet. Their participants reviewed materials that presented rape as very controllable, somewhat controllable, or not at all controllable. Women who read the uncontrollable version reported greater levels of anxiety about rape than did women who read the other versions. However, the researchers did not actually intend to produce a new rape-prevention pamphlet. Participants in this research were deceived regarding the true purpose of the study: to investigate how perceived controllability of rape influences women's perceptions of vulnerability to rape. Under what conditions should researchers be allowed to deceive research participants?

Deception is just one of many ethical issues that researchers must confront. As yet another illustration of ethical problems, we mentioned that animal subjects sometimes are used to help understand human psychopathology. This may mean exposing animal subjects to stressful and even painful conditions. Again, we must ask about the ethical issues involved with this type of research. Under what conditions should research with animal subjects be permitted? The list of ethical questions raised by psychological research is a lengthy one. Thus, *it is of the utmost importance that you become familiar with the APA ethical principles and their application at an early stage in your research career, and that you participate (as research participant, assistant, or principal investigator) only in research that meets the highest standards of scientific integrity.*

Getting Started Doing Research

— What to study?
— How do I find a hypothesis to test?
— Is my research question a good one?

There are many decisions that must be made before beginning to do research in psychology. The first one, of course, is, what to study. In what area of psychology do I wish to do research? Many students approach the field of psychology with interests in psychopathology and issues associated with mental illness. Others are intrigued with the puzzles surrounding human cognition, such as memory, problem solving, and decision making. Still others are interested in problems of developmental and social psychology. Psychology provides a smorgasbord of research possibilities to explore, as is illustrated by the literally hundreds of scientific journals that publish the results of psychological research. You can quickly find information about the many research areas within psychology by reviewing the contents of a standard introductory psychology textbook. More specific information can be found, of course, in the many classes offered by the psychology department of your college or university, such as abnormal psychology, cognitive psychology, and social psychology.

The next decision is a bit harder. How do I come up with a research hypothesis? McGuire (1997) believes that the ability to generate testable hypotheses can be described and taught to students. He identified 49 simple rules ("heuristics") for generating a hypothesis to be tested scientifically. We can not review all 49 suggestions here, but we can give you some insight into McGuire's thinking about scientific thinking by listing some of these heuristics. He suggests, for example, that we might generate a hypothesis for a research study by:

— thinking about deviations (oddities, exceptions) from a general trend or principle;
— imagining how *we* would behave in a task or faced with a specific problem;
— considering similar problems whose solution is known;
— making sustained, deliberate observations of a person or phenomenon (e.g., performing a case study);
— considering the opposite direction of plausible causality for a phenomenon;
— generating counterexamples for an obvious conclusion about behavior;
— exploring alternative labels for phenomena;
— "jolting one's conceptualizing out of its usual ruts" by shifting attention to the opposite pole of the problem or expressing a question in different modalities (e.g., pictures, words, symbols);
— coming up with multiple explanations for a given finding;
— thinking about how people need to function in a given environment;
— borrowing ideas or theories from other disciplines;
— looking for irregularities in previous results and attempting to reconcile conflicting outcomes;
— trying new modes of data collection such as performing qualitative analyses of a quantitative finding or carrying out new statistical analyses of old data.

No matter how or where you begin to develop a hypothesis for your research, at some point you will find yourself needing to peruse the published literature of psychological research. There are several reasons why you must search the psychology literature before beginning to do research. One obvious reason is that the answer to your research question may already be there. Someone else may have entertained the same question and provided an answer, or at least a partial one. It is highly unlikely that you will not find some research finding that is related to the research question you have chosen to answer. This should not discourage you; in fact, it should be a source of satisfaction. Finding that other people have done research on the same or similar idea will give you the "in" that is necessary to build the science of psychology. Doing research without a careful examination of what is already known may be interesting or fun (it certainly may be easy); perhaps you could call it a "hobby," but we can't call it science. Science is a cumulative affair: Current research builds on previous research. Once you have identified a body of literature related to your research idea, your reading may lead you to discover, as McGuire (1997) suggested, inconsistencies or contradictions in the published research. You may also find that the research findings are limited in terms of the nature of the participants studied or the circumstances under which the research was done, or that there is a psychological theory in need of testing. Having made such a discovery, you have found a solid research lead, a path to follow.

When reading the psychological literature and thinking about possible research questions, you might also consider how the results of psychological studies might be applied to societal problems. Later we discuss differences between basic and applied research (see especially Chapter 10). Suffice it to say that applied research typically focuses on improving the human condition. As you learn how to do research in psychology, you may want to consider ways this knowledge can be used to generate questions that when answered will make humankind just a little better off.

Searching the psychology literature is not the tedious task that it once was; computer-aided literature searches, including use of the Internet, have made identifying related research a relatively easy, even exciting task. In Appendix C of this book we outline how to search the psychology literature, including ways to use computer-based guides.

FIGURE 1.6 Computers have made reviewing psychological literature a relatively easy task.

Finally, as Sternberg (1997) points out, choosing a question to investigate should not be taken lightly. Some questions are simply not worth asking because their answers offer no hope of advancing the science of psychology. The questions are, in a word, meaningless, or at best, trivial. "How do I know if my research question is a good one?" Sternberg (1997) suggests that students new to the field of psychological research ask themselves several questions before deciding they have a good research question:

— Why might this question be scientifically important?
— What is the scope of this question?
— What are likely outcomes if I carry out this research project?
— To what extent will psychological science be advanced by knowing the answer to this question?
— Why would anyone be interested in the results obtained by asking this question?

As you begin the research process, finding answers to these questions may require guidance from research advisors and others who have successfully answered them in their own research. We also hope that your ability to answer these questions will be enhanced as you learn more about theory and research in psychology and as you read about the many examples of interesting and meaningful psychological research that we describe in this book.

Of course, identifying a research problem doesn't necessarily tell you *how* to do the research. What is it exactly that you want to know? Answering this question will mean that you must make other decisions as well. Should I do a qualitative or quantitative research study? What is the nature of the variables I wish to investigate? How do I find reliable and valid measures of behavior? What is the research method best suited to my research question? What kinds of statistical analyses will be needed? Do the methods I choose meet accepted moral and ethical standards?

A major goal of this book is to provide you with the information to help make these decisions and many others associated with the research process.

The Organization of This Book

This book provides an introduction to the way in which the scientific method is applied in psychology. As you are probably aware, the scope of psychology is quite large, encompassing many areas of study. Moreover, no single research methodology can be said to answer all the questions raised in a particular area. Thus, the best approach to answering a question about behavior or mental processes is frequently a *multimethod approach*—that is, searching for an answer using different research methodologies.

The organization of the book follows loosely the goals for the scientific method. Following a discussion of ethical issues in research (Chapter 2), we introduce you to what are called *descriptive methods*. Naturalistic observation, for instance, is an important tool of psychologists who want to describe behavior in a natural context (see Chapter 3). Questionnaires are among the most commonly used instruments in psychology and frequently are the basis for making *predictions* about behavior. The nature of questionnaires and their use in survey research are discussed in Chapter 4. Several less commonly used measures of behavior—those based on the examination of archival records and those derived from the study of physical traces—are treated in Chapter 5. In the third part of the book we deal with *experimental approaches* to the study of behavior. The emphasis is on experiments done with groups of participants (see Chapters 6–8). Experimental approaches are aimed chiefly at discovering cause-and-effect relationships. However, experimental methods used in the laboratory are not necessarily the same as those used outside the laboratory.

Finally, in the fourth section of the book we introduce you to the methods employed in *applied research*. In Chapter 9 there is a discussion of experiments conducted with small numbers of research participants, in fact, with single individuals ($N = 1$). In that chapter we also discuss the important case study method. The last chapter deals with quasi-experimental designs and with the important topic of program evaluation, which is concerned with assessing the effects of "treatments" applied in natural settings (see Chapter 10).

SUMMARY

Psychology's official beginning is marked by the establishment, in 1879, of a formal psychology laboratory in Leipzig, Germany, under the direction of Wilhelm Wundt. With this beginning came the first applications of the scientific method to problems of psychology. As an approach to knowledge, the scientific method is characterized by a reliance on empirical procedures, rather than intuition, and by an attempt to control the investigation of those factors believed responsible for a phenomenon. Those factors that are systematically controlled in an attempt to determine their effect on behavior are called independent variables. The measures of behavior used to assess the effect (if any) of the independent variable are called dependent variables.

Scientists seek to report results in an unbiased and objective manner. This goal is enhanced by giving operational meaning to concepts. Scientists also seek to measure phenomena as accurately and precisely as possible. Measurement involves both physical and psychological measurement. Scientists seek both validity and reliability of these measures.

Hypotheses are tentative explanations of events. To be useful to the scientist, however, hypotheses must be testable. Hypotheses that lack adequate definition, that are circular, or that appeal to ideas or forces outside the province of science are not testable. Hypotheses are often derived from theories. More than anything else, scientists are skeptical. A skeptical attitude is not always found among nonscientists, who may rush to accept "new discoveries" and extraordinary claims.

The goals of the scientific method are description, prediction, and understanding. Both quantitative and qualitative research are used to describe behavior. Observation is the principal basis of scientific description. When two measures correlate, we can predict the value of one measure by knowing the value of the other. Understanding is achieved when the causes of a phenomenon are discovered. This requires that evidence be provided for covariation of events, that a time-order relationship exists, and that alternative causes be eliminated. When two potentially effective variables covary such that the independent effect of each variable on behavior cannot be determined, we say that our research is confounded. Confounding must be avoided if we wish to produce a study with internal validity. The external validity of a study involves the extent to which research results can be generalized to different populations, settings, and conditions.

Scientific theory construction and testing provide the bases for a scientific approach to psychology. Theories have the important function of guiding research and organizing empirical knowledge. Finally, many ethical questions are raised by psychological research; it is important that the science of psychology be carried out according to the highest standards of scientific integrity. Getting started doing psychological research requires us to make several important decisions, including those about what topic to investigate, what is the specific question I wish to answer, and is my question a good one? Finally, we must decide exactly how to *do* the research.

Key Concepts

scientific method
empirical approach
control
independent variable
individual differences variable
dependent variable
operational definition
validity
reliability
hypothesis
correlation
causal inference
confounding
internal validity
external validity
theory

Review Questions

1. For each of the following characteristics, indicate how the scientific approach differs from nonscientific (everyday) approaches to knowledge: general approach, observation, reporting, concepts, instruments, measurement, hypotheses, and attitudes.
2. What is the major advantage of using operational definitions in psychology? What disadvantages are there in using operational definitions?
3. Distinguish between the accuracy of a measuring instrument and its precision.
4. What distinguishes physical measurement from psychological measurement?
5. What three shortcomings often keep hypotheses from being testable?
6. Why do behavioral scientists always seek to maintain a skeptical attitude?
7. What are the three goals of the scientific method?
8. What do we mean when we say that the nomothetic approach is used in science? How is the nomothetic approach different from the idiographic approach?

9 Provide an example both of quantitative research and of qualitative research.
10 What three conditions must be met if one event is to be considered the cause of another?
11 Distinguish between the internal validity and the external validity of a research study.
12 Describe how a scientific theory can be used to guide and organize empirical research.
13 Identify at least three characteristics used to evaluate a theory.
14 Explain the use of intervening variables in theory construction. What important quality should an intervening variable have?

CHALLENGE QUESTIONS

1 In each of the following descriptions of research studies, you are to identify the independent variable(s). You should also be able to identify at least one dependent variable in each study.

A A psychologist was interested in the effect of food deprivation on motor activity. She assigned each of 60 rats to one of four conditions differing in the length of time for which the animals were deprived of food: 0 hours, 8 hours, 16 hours, 24 hours. She then measured the amount of time the animals spent in the activity wheel in their cages.

B A physical education instructor was interested in specifying the changes in motor coordination with increasing age in young children. He selected six groups of children and gave each child a test of motor coordination. The groups of children differed in age—that is, one group was made up of all 5-year-olds, the next group was all 6-year-olds, and so on up to the sixth group, which was all 10-year-olds.

C A developmental psychologist was interested in the amount of verbal behavior very young children displayed depending on who else was present. The study he did involved selecting children who were either 2, 3, or 4 years old. These children were observed in a laboratory setting for a 30-minute period. Half of the children of each age were assigned to a condition in which an adult was present with the child during the session. The other half of the children were assigned to a condition in which another young child was present during the session with the child being observed. The psychologist measured the number, duration, and complexity of the verbal utterances of each observed child.

2 In the following description the independent variable of interest is confounded with a potentially relevant independent variable. Identify the confounding variable and explain clearly how the confounding occurred. Also state exactly what conclusion can be supported on the basis of the evidence presented. Finally, suggest ways in which the study could be done so that it would be internally valid.

A physiological psychologist developed a drug that she thought would revolutionize the world of horse racing. She named the drug Speedo, and it was her contention that this drug would lead horses to run much faster than they do now. (Forget, for the sake of this problem, that it is illegal to give drugs to racehorses.) She selected two groups of horses and gave one of the groups injections of Speedo once a week for 4 weeks. Because Speedo was known to have some negative effects on the horses' digestive systems, those horses given the Speedo had to be placed on a special high-protein diet. Those horses not given the Speedo were maintained on their regular diet. After the 4-week period, all the horses were timed in a 2-mile race and the mean times for the horses given Speedo were significantly faster than the mean times for those not given Speedo. The psychologist concluded that her drug was effective.

3 Two physicians did a study to try to determine why people who are allergic to cats still keep them as pets. They asked 67 patients (22 male and 45 female) with an average age of 40 years to complete a questionnaire concerning the nature of their attachment to their cat. Even though 11 people had a history of emergency room visits following exposure to cats, 38 said they would definitely replace their cat if it died, and an additional 16 reported they would have difficulty avoiding a new acquisition. Having someone to love and companionship were the most commonly selected reasons for having cats. The physicians concluded that cat ownership meets strong psychological needs in allergic patients. What comparison information is needed before reaching the conclusion that the psychological reasons for owning a cat are peculiar to allergic patients?

4 When presented with only correlational evidence, the investigator can only hypothesize about possible causal factors underlying a relationship between variables. A stage wherein the investigator thinks about possible causal factors and "tries them out" as explanations, perhaps during discussions with other researchers, is often a preliminary step to doing research that may provide evidence for the factors responsible for the reported relationship. For each of the following reports of covariation, identify a possible causal factor—that is, speculate on *why* these events are correlated.

A A study of nearly 5,000 Swedish women (20 to 44 years old) revealed that couples who live together before marriage have an 80% higher probability of getting divorced than do married couples who do not live together before marriage.

B Married women are found to suffer from depression more than men. However, the greater the marital compatibility, the less the risk of depression among women.

C In the late 1980s the National Institutes of Health sponsored a conference to assess available data regarding the healing power of pets. The idea is that having pets is good for you. One investigator reported that among 92 coronary patients, pet owners were more likely to be alive 1 year after discharge from a coronary heart unit than were people who do not own a pet.

Answer to Challenge Question 1

A IV: hours of food deprivation with four levels; DV: time (in minutes) animals spent in activity wheel

B IV: age of children with six levels (distinguish selection of levels from manipulation in part A); DV: scores on test of motor coordination

C IV: age at three levels and additional person present with two levels (again, distinguish selection and manipulation); DV: number, duration, and complexity of child's verbal utterances