CHAPTER TWO

The Scientific Method

CHAPTER OUTLINE

Scientific and Everyday Approaches to Knowledge

- The scientific method is empirical and requires systematic, controlled observation.
- Scientists gain the greatest control when they conduct an experiment; in an experiment, researchers manipulate independent 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 constructs 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.

For over 100 years the scientific method has been the basis for investigation in the discipline of psychology. The scientific method does not require a particular type of equipment, nor is it associated with a particular procedure or technique. As first described in Chapter 1, the scientific method refers to the ways in which scientists ask questions and the logic and methods used to gain answers. There are many fruitful approaches to gaining knowledge about ourselves and our world, such as philosophy, theology, literature, art, and other disciplines. The scientific method is distinguishable from the other approaches, but all of them share the same goal—seeking the truth. One of the best ways to understand the scientific method as a means of seeking truth is to distinguish it from our "everyday" ways of knowing. Just as a telescope and a microscope extend our everyday abilities to see, the scientific method extends our everyday ways of knowing.

Several major differences between scientific and our everyday ways of knowing are outlined in Table 2.1. Collectively, the characteristics listed under "Scientific" define the scientific method. The distinctions made in Table 2.1

TABLE 2.1 CHARACTERISTICS OF SCIENTIFIC AND NONSCIENTIFIC (EVERYDAY) APPROACHES TO KNOWLEDGE*

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

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

highlight differences between the ways of thinking that characterize a scientist's approach to knowledge and the informal and casual approach that often characterizes our everyday thinking. These distinctions are summarized in the following pages.

General Approach

We described in Chapter 1 that in order to think like a researcher you must be skeptical. Psychological scientists are cautious about accepting claims about behavior and mental processes, and they critically evaluate the evidence before accepting any claims. In our everyday ways of thinking, however, we often accept evidence and claims with little or no evaluation of the evidence. In general, we make many of our everyday judgments 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. The many everyday inferences and conclusions we reach intuitively are the product of insight and of what we quickly perceive as true based on our personal experiences. Although intuition can be valuable when we have little other information, intuition is not always correct. Consider, for example, what intuition might suggest regarding ratings of video games, movies, and television programs for violent and sexual content. Intuition might suggest that ratings are effective tools for preventing exposure to violent content. In fact, just the opposite may take place! Research indicates that these ratings can entice adolescent viewers to watch the violent and sexy programs—what Bushman and Cantor (2003) called a "forbidden-fruit effect." Thus, rather than limiting exposure to violent and sexual content, ratings may *increase* exposure because "ratings may serve as a convenient way to find such content" (p. 138).

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 cognitive biases, or because we neglect to weigh available evidence appropriately (Kahneman & Tversky, 1973; Tversky & Kahneman, 1974). Daniel Kahneman won the Nobel Prize in 2002 for his research on how cognitive biases influence people's economic choices. One type of cognitive bias, called illusory correlation, is our tendency to perceive a relationship between events when none exists. Susskind (2003) showed that children are susceptible to this bias when they make judgments about men's and women's behaviors. Children in 2nd grade and 4th grade were shown many pictures of men and women performing stereotypical (e.g., a woman knitting), counterstereotypical (e.g., a man knitting), and neutral behaviors (e.g., a woman or a man reading a book). The children's task was to estimate how frequently they saw each picture. The results indicated that children overestimated the number of times they saw pictures displaying stereotypical behavior. By responding in this way, the children showed that they were susceptible to an illusory correlation. That is, their expectations that men and women behave in stereotypical ways led the children to believe that these types of pictures were displayed more often than they were. 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.

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 the scientist's intuitive hunches as to what direction to take. Eventually, however, the scientist strives to be guided by the empirical evidence that direct observation and experimentation provide.

Observation

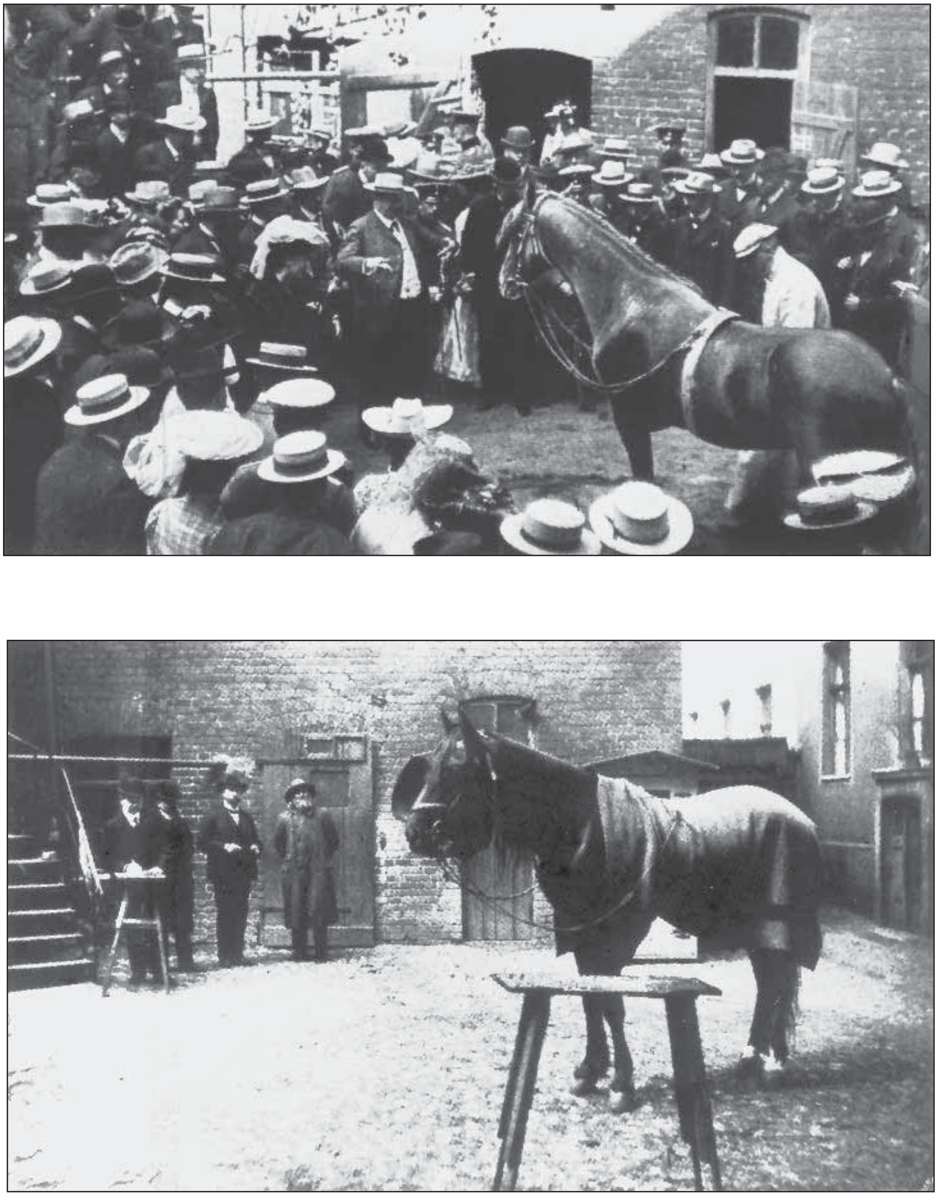
We can learn a great deal about behavior by simply observing the actions of others. However, everyday observations are not always made carefully or systematically. Most people do not attempt to control or eliminate factors that might influence the events they are observing. As a result, we often make incorrect conclusions based on our casual observations. 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?"), 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' performances, and hundreds of people came to view this amazing horse (Figure 2.1). In 1904 a scientific commission was established with the goal of discovering the basis for Hans' abilities. Much to his owner's dismay, the scientists observed that Hans was not clever in two situations. 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. What did the scientists observe? They discovered that Hans was responding to the questioner's subtle movements. 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 as he tapped his forefoot or pointed. Thus, it seems that Hans was a better observer than many of the people who observed him!

Key Concept

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 2.1). By using

FIGURE 2.1 Top: Clever Hans performing before onlookers. Bottom: Hans being tested under more controlled conditions when Hans could not see the questioner.



controlled observation, scientists gain a clearer picture of the factors that produce a phenomenon. The careful and systematic observation of Clever Hans is one example of the control used by scientists to gain understanding about behavior. Box 2.1 describes an example of how the story of Clever Hans from over 100 years ago informs scientists even today.

BOX 2.1

CAN DOGS DETECT CANCER? ONLY THE NOSE KNOWS

Research on methods to detect cancer took an interesting turn in 2004 when investigators reported the results of a study in the *British Medical Journal* demonstrating that dogs trained to smell urine samples successfully detected patients' bladder cancer at rates greater than chance (Willis et al., 2004). This research followed up many anecdotal reports in which dog owners described their pets as suddenly overprotective or obsessed with skin lesions prior to the owners' being diagnosed with cancer. Interest in the story was so great that similar demonstrations were conducted on television programs such as *60 Minutes.*

Skeptics, however, cited the example of Clever Hans to challenge the findings, arguing that the dogs relied on researchers' subtle cues in order to discriminate samples taken from cancer *vs.* control patients. Proponents of the study insisted that the researchers and observers were blind to the true status of the samples so could not be cuing the dogs. Researchers in this new area of cancer detection have applied for research funding to conduct more experiments. We now await the results of these rigorous studies to tell us whether dogs can, in fact, detect cancer.

Key Concepts

Scientists gain the greatest control when they conduct an experiment. In an **experiment,** scientists manipulate one or more factors and observe the effects of this manipulation on behavior. The factors that the researcher controls or manipulates in order to determine their effect on behavior are called the **independent variables.**[1] 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. For example, if we wanted to study the effect of drinking alcohol on the ability to process complex information quickly

[1]Sometimes the levels of the independent variable are *selected* by a researcher rather than manipulated. An *individual differences variable* is a characteristic or trait that varies across individuals; for example, sex of the participants (male, female) is an individual differences variable. When researchers investigate whether behavior differs according to participants' sex, they select men and women and examine this factor as an individual differences variable. As we will see in Chapter 7, there are important differences between manipulated and selected independent variables.

and accurately, the independent variable would be the presence or absence of alcohol in a drink. Participants in the experimental condition would receive alcohol, while participants in the control condition would receive the same drink without alcohol. After manipulating this independent variable, the researcher might ask participants to play a complicated video game to see whether they are able to process complex information.

Key Concept

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 the difficult video game. The number of errors, then, would be the dependent variable.

Scientists seek to determine whether any differences in their observations of the dependent variable are caused by the different conditions of the independent variable. In our example, this would mean that a difference in errors when playing the video game is caused by the different independent variable conditions—whether alcohol is present or absent. To form this clear conclusion, however, scientists must use proper control techniques. Each chapter of this book will emphasize how researchers use control techniques to study behavior and the mind.

Reporting

Suppose you ask someone to tell you about a class you missed. You probably want an accurate report of what happened in class. Or perhaps you missed a party at which two of your friends had a heated argument, and you want to hear from someone what happened. As you might imagine, personal biases and subjective impressions often enter into everyday reports that we receive. When you ask others to describe an event, you are likely to receive details of the event (not always correct) along with their personal impressions. You may also find that the details reported to you are not the ones 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 2.2. How would you describe to someone what you see there? One way to describe this scene is to say that three people are running along a path. 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 you have seen and not just reporting what you have observed. The description of three people running would be preferred in a scientific report.

FIGURE 2.2 How would you describe this scene?

This distinction between description and inference in reporting can be carried to extremes. For example, describing what is shown in Figure 2.2 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 4.

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. 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.

STRETCHING EXERCISE

In this exercise we ask you to respond to the questions that follow this brief description of a research report.

A relatively new area of psychology called "positive psychology" focuses on positive emotion, positive character traits, and positive institutions; the goal of research in positive psychology is to identify ways to foster well-being and happiness (Seligman, Steen, Park, & Peterson, 2005). One area of research focuses on *gratitude,* the positive emotion people feel when they are given something of value by another person (Bartlett & DeSteno, 2006). Some research suggests that people who feel gratitude are more likely to act prosocially—that is, to behave in ways that benefit others.

Bartlett and DeSteno (2006) tested the relationship between gratitude and participants' likelihood of helping another person in an experiment involving *confederates* (people working with the experimenter to create an experimental situation; see Chapter 4). Each participant first teamed up with a confederate to complete a long, boring task involving hand-eye coordination. Afterward, for one third of the participants their computer screen was designed to go blank and they were instructed they would need to complete the task again. The confederate, however, induced an emotion of gratitude by fixing the problem, saving the participant from having to redo the task. The situation differed for the other participants. After finishing the task, one third of the participants watched an amusing video with the confederate (positive emotion) and the final one third of the participants had a brief verbal exchange with the confederate (neutral emotion). After completing some questionnaires, the confederate asked each participant to fill out a lengthy survey for one of her classes as a favor. Bartlett and DeSteno found that participants in the gratitude condition spent more time working on the survey ($M = 20.94$ minutes) than participants in the positive emotion ($M = 12.11$ min) and neutral emotion ($M = 14.49$ min) conditions.

1. Identify the independent variable (including its levels) and the dependent variable in this study.
2. How could the researchers determine that it was *gratitude*, not simply feeling positive emotions, that increased participants' willingness to help the confederate?

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.

In everyday conversation we often get by without having to worry about how we define a concept. Many words, for instance, are commonly used and apparently understood even though neither party in the conversation knows exactly what the words mean. That is, 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 learned. You might try arguing a point of view opposite to theirs just for the fun of it. After discussing 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, even after debating its origins. 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 mostly inherited or mostly learned, we need to have an exact definition that all parties involved can accept.

Key Concepts

The study of "concepts" is so important in psychological science that researchers refer to concepts by a special name: constructs. A **construct** is a concept or idea; examples of psychological constructs include intelligence, depression, aggression, and memory. One way in which a scientist gives meaning to a construct is by defining it operationally. An **operational definition** explains a concept solely in terms of the observable procedures used to produce and measure it. Intelligence, for instance, can be defined operationally by using 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

FIGURE 2.3 If balancing a ball on your nose is an operational definition of intelligence, would seals be considered more intelligent than humans?

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 construct 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 construct 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, a major strength of using operational definitions is that they help to clarify communication among scientists about their constructs. This strength is assumed to outweigh the limitations.

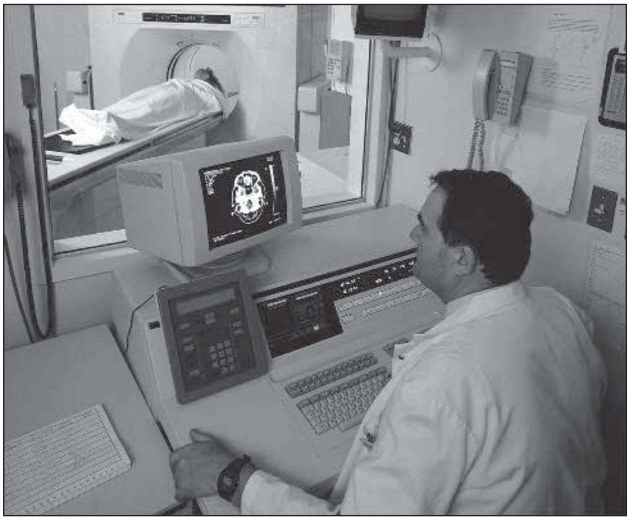
Instruments

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 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 not ill when we are. The accuracy of an instrument is determined by *calibrating* it, or checking it with another instrument known to be true. The accuracy of speedometers can be checked using a combination of observations of roadside distance markers and measuring the elapsed time using 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 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. Wilhelm Wundt established a formal psychology laboratory in 1879, an event that marks the official beginning of scientific psychology. Wundt used a reaction-time apparatus to measure the time required for cognitive processing. You can be assured that the precision, and

FIGURE 2.4 Scientific instruments used in psychology have improved dramatically in their precision and accuracy.



even the accuracy, of instruments of this kind have improved significantly since then. 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. So, too, are the rating scales used by human observers. 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 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

In order to investigate events and phenomena, scientists use instruments to obtain measurements. Measurements provide the record of the careful and controlled observations that characterize the scientific method. One type of scientific measurement, *physical measurement,* involves dimensions for which there is an agreed-upon standard and an instrument for doing the measuring. For example, length is a dimension that can be scaled with physical measurement, and there are agreed-upon standards for units of length (e.g., inches, meters). Similarly, units of weight and time represent physical measurement.

Although researchers in psychology use physical measurement, most of the dimensions measured in psychological research do not involve physical measurement. Rulers do not exist for measuring psychological constructs such as beauty, aggression, or intelligence. For these dimensions we must use a second type of measurement—*psychological measurement*. In a sense, the human observer is the instrument for psychological measurement. More specifically, agreement among a 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.

Key Concept

Just as it is important that instruments be accurate and as precise as possible, 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 construct 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 question whether this is a valid measure of intelligence. 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 construct. 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 on measures such as size of vocabulary, reasoning ability, and other accepted measures of intelligence.

Key Concept

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 (Medvec, 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).

The validity and reliability of measurements are central issues in psychological research. You will encounter various ways in which researchers determine reliability and validity as we introduce you to different research methods.

Hypotheses

A hypothesis is a tentative explanation for something. Hypotheses frequently attempt to answer the questions "How?" and "Why?" At one level, a hypothesis may simply suggest how particular variables are related. For example, in our popular culture we frequently associate white or brightness with "good" and black or darkness with "bad" (Meier, Robinson, & Clore, 2004). In the movie *Star Wars*, for instance, Luke Skywalker and Princess Leia were dressed in white and Darth Vader was completely in black. Across many religions (e.g., Buddhism, Christianity, Hinduism, Islam, Zoroastrianism) an association exists between light and God or goodness, and between darkness and Satan or evil. Whether something is considered good or bad is referred to as an *affective* judgment. On the other hand, our experience of brightness (and darkness) is a sensory perception. In their research, Meier and his colleagues hypothesized that the association between affective judgments and sensory perceptions of brightness is automatic—that is, people automatically judge brighter objects as good and darker objects as bad.

To test their hypothesis, Meier and his associates (2004) asked participants in a series of experiments to judge whether 100 words presented on a computer screen were *negative* or *positive*. Fifty of the words were previously rated as reflecting positive affect (e.g., *candy, love, pretty, sleep*), and 50 of the words represented negative affect (e.g., *bitter, cancer, devil, rude*). The researchers manipulated whether the words were presented in a bright font or a dark font. Their results indicated that when the affect and the brightness of the word conflicted (e.g., *love* presented in a dark font), participants took longer and made more errors when judging whether the word was positive or negative, compared to when the words "matched" the associated brightness (e.g., *love* presented in a bright font).

At a more theoretical level, a hypothesis may offer a reason (the "why") for the way particular variables are related. For example, Meier and his colleagues (2004) considered theories that suggest the human brain has developed in a way that makes conceptual thinking, such as affective judgments, automatically tied to physical perception (e.g., Barsalou, 1999; Lakoff & Johnson, 1999). Based on these theories, the researchers suggested that people cannot judge the affect of a word (or any other object) without first automatically considering its physical features, such as brightness. In their experiment, when the brightness conflicted with the correct affective judgment, additional processing (i.e., time, thought) was required for people to override their automatic association and to make the correct judgment about whether the word was negative or positive.

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 that distinguishes casual, everyday hypotheses from scientific hypotheses is *testability*. If a hypothesis cannot be tested, it is not useful to science (Marx, 1963). Three types of hypotheses fail to pass the "testability test." A hypothesis is not testable when its constructs are not adequately defined, when the hypothesis is circular, or when the hypothesis appeals to ideas not recognized by science.

Hypotheses are not testable if the concepts to which they refer are not adequately defined. Consider a hypothesis saying that a would-be assassin shot a U.S. president or other prominent figure because he was mentally disturbed. This hypothesis would not be testable 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 because an accepted operational definition is often not available for these concepts. 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. A circular hypothesis occurs when an event itself is used as the explanation of the event (Kimble, 1989, p. 495). As an illustration, consider the statement that an "eight-year-old boy is distractable in school and having trouble reading because he has an attention deficit disorder." An attention deficit disorder is defined by the inability to pay attention. Thus, the statement simply says that the boy doesn't pay attention because he doesn't pay attention—that's a circular hypothesis.

A hypothesis also may be untestable if it appeals to ideas or forces that are not recognized by science. 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.

Goals of the Scientific Method

- The scientific method is intended to meet four goals: description, prediction, explanation, and application.

In the first part of this chapter, we examined the ways in which our everyday ways of thinking differ from the scientific method. In general, the scientific method is characterized by an empirical approach, systematic and controlled observation, unbiased and objective reporting, clear operational definitions of constructs, accurate and precise instruments, valid and reliable measures, and testable hypotheses. In this next section, we examine goals of the scientific method. Psychologists use the scientific method to meet four research goals: description, prediction, explanation, and application.

Description

- Psychologists seek to describe events and relationships between variables; most often, researchers use the nomothetic approach and quantitative analysis.

Description refers to the procedures researchers use to define, classify, catalogue, or categorize events and their relationships. Clinical research, for instance, provides practitioners with 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., Text Revision, 2000), also known as DSM-IV-TR (see Figure 2.5). Consider, as one example, the criteria used to define the disorder labeled dissociative fugue (formerly psychogenic fugue).

FIGURE 2.5 Clinicians classify mental disorders according to the criteria found in the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders.*

Diagnostic Criteria for Dissociative Fugue

A The predominant disturbance is sudden, unexpected travel away from home or one's customary place of work, 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).

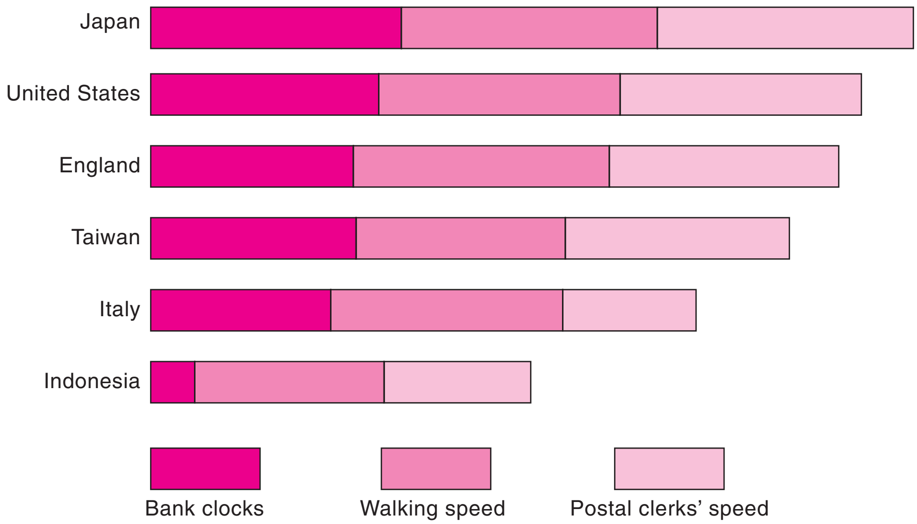
D The symptoms cause clinically significant stress or impairment in social, occupational, or other important areas of functioning. (DSM-IV-TR, 2000, p. 526)

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." Researchers also seek to provide clinicians with descriptions of the prevalence of a mental disorder as well as the relationship between the presence of various symptoms and other variables such as gender and age. According to the DSM-IV-TR (2000), 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. 524).

Science in general and psychology in particular develop descriptions of phenomena using the *nomothetic approach*. Using the nomothetic approach, psychologists try to establish broad generalizations and general laws that apply to a diverse population. To accomplish this goal, psychological studies most often involve large numbers of participants. Researchers seek to describe the "average," or typical, performance of a group. This average may or may not describe the performance of any one individual in the group.

For example, Levine (1990) described the "pace of life" in various cultures and countries of the world by noting the accuracy of outdoor bank clocks and by timing the walking speed of pedestrians over a distance of 100 feet. The results of this study are shown in Figure 2.6. The citizens of Japan exhibited, overall, the fastest pace of life with U.S. citizens second. The citizens of Indonesia were the slowest. Not all citizens of Japan or the United States, however, are on the fast track. In fact, Levine (1990) and his colleagues found wide differences in the pace of life among various cities within the United States 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 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.

FIGURE 2.6 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.)



Researchers who use the nomothetic approach appreciate that there are important differences among individuals; they seek, however, to emphasize the similarities rather than the 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 (e.g., rewards, punishments). Researchers merely seek to describe what organisms are like in general on the basis of the average performance of a group of different organisms.

Some psychologists, notably Gordon Allport (1961), argue that the nomothetic approach is inadequate—unique individuals cannot be described by an average value. Researchers who use the *idiographic approach* study the individual rather than groups. These researchers believe that although individuals behave in ways that conform to general laws or principles, the uniqueness of individuals must also be described. A major form of idiographic research is the case study method, which we will describe in Chapter 10.

Depending on their research question, researchers decide whether to describe groups of individuals or one individual's behavior. 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 in which the findings are mainly the product of statistical summary and analysis. *Qualitative research* produces verbal summaries of research findings with few statistical summaries or analysis. Just as psychological research is more frequently nomothetic than idiographic, it is also more typically quantitative than qualitative.

Qualitative research is used extensively by sociologists and anthropologists (see, for example, Seale, 1999). 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). Central to qualitative research is that investigators ask participants to describe their experiences in ways that are meaningful to *them,* rather than asking participants to use categories and dimensions established by theorists and previous research (Kidd, 2002). This qualitative approach was used by Kidd and Kral (2002) to gain insight into the experiences of 29 Toronto street youth (ages 17–24). A focus of the interviews concerned experiences with suicide. The majority (76%) of those interviewed reported a history of attempted suicide, and analysis of their narratives revealed that suicidal experiences were linked especially to feelings of isolation, rejection/betrayal, low self-worth, and prostitution. Importantly, the researchers reported that their analyses revealed several topics associated with suicidal experiences not identified in previous research involving street youth. Namely, "loss of control, assault during prostituted sex, drug abuse as a 'slow suicide,' and breakups in intimate relationships" were related to these youths' suicidal experiences (p. 411). Other examples of qualitative research are found in Chapter 4 when we discuss narrative records of observed behavior; case studies described in Chapter 10 also are a form of qualitative research.

Prediction

- Correlational relationships allow psychologists to predict behavior or events, but do not allow psychologists to infer what causes these relationships.

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 also is helpful in both the treatment and 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, scores on the Graduate Record Examination (GRE), as well as undergraduate grade point average (GPA), can be used to predict how well a student will do in graduate school. Sternberg and Williams (1997) did find that GRE scores predicted fairly well the first-year grades of graduate students at their institution. They also found, however, that 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 predictive validity (i.e., accuracy of prediction) of the GRE, which is widely regarded as a predictor of students' later professional development (see, for example, "Comment" section of *American Psychologist*, 1998, 53, 566–577).

Key Concept

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." For example, stress and illness are known to be correlated; the more stressful life events people experience, the more likely they are to experience physical illnesses.

Consider a measure with which you likely have had some experience, namely, teacher/course evaluations in classes 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, likable). 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. That is, more positive course evaluations of teachers were associated with higher ratings for their videotaped behavior; similarly, more negative course evaluations were associated with lower ratings of videotaped behavior. 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., likableness).

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.

Interestingly, 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). Ambady and Rosenthal (1993) proposed an explanation for their correlation between teacher evaluations by students not enrolled in the class and by students enrolled in the class. They suggested that people are "attuned" to picking up information about a person's affect quickly because this information is important (adaptive) in real-life decision making. Without additional information, however, the proposed explanations for these two phenomena are speculative.

Explanation

- 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.
- Researchers seek to generalize a study's findings to describe different populations, settings, and conditions.

Although description and prediction are important goals in science, they are only the first steps in our ability to explain and understand a phenomenon. Explanation is the third goal of the scientific method. We understand and can explain a phenomenon when we can identify its causes. Researchers typically conduct *experiments* to identify the causes of a phenomenon. Experimental research differs from descriptive and predictive (correlational) research because of the high degree of control scientists seek in experiments. Recall that when researchers control a situation, they manipulate independent variables one at a time to determine their effect on the dependent variable—the phenomenon of interest. By conducting controlled experiments, psychologists infer what causes a phenomenon; they make a causal inference. Because experiments are very important to psychologists' efforts to form causal inferences, we have dedicated Chapters 7, 8, and 9 to a detailed discussion of the experimental method.

Key Concept

Scientists set three important conditions for making a **causal inference:** *covariation of events, a time-order relationship,* and *the elimination of plausible alternative causes.* A simple illustration will help you to understand these three conditions. Suppose you hit your head on a door and experience a headache; presumably you would *infer* that hitting your head *caused* the headache. The first condition for causal inference is covariation of events. If one event is the cause of another, the two events must vary together; that is, when one changes, the other must also change. In our illustration, the event of changing your head position from upright to hitting against the door must covary with experience of no headache to the experience of a headache.

The second condition for a causal inference is a *time-order relationship* (also known as contingency). The presumed cause (hitting your head) must occur before the presumed effect (headache). If the headache began before you hit your head, you wouldn't infer that hitting your head caused the headache. In other words, the headache was contingent on you hitting your head first. Finally, causal explanations are accepted only when other possible causes of the effect have been ruled out—when *plausible alternative causes have been eliminated.* In our illustration, this means that to make the causal inference that hitting your head caused the headache, you would have to consider and rule out other possible causes of your headache (such as reading a difficult textbook).

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, it has been 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. It is also the case, 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 a teacher wished to demonstrate that so-called "active learning strategies" (e.g., debates, group presentations) 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? Perhaps another group of students taught using a different approach might learn the same amount. Before the teacher could claim that active learning stategies *caused* the performance she observed, she would have to compare this method with some other reasonable approach. That is, 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. When a controlled experiment is done, a bonus comes along when the independent and dependent variables covary. The time-order condition for a causal inference is met because the researcher manipulates the independent variable (e.g., teaching method) and *subsequently* measures the differences between conditions on the dependent variable (e.g., a measure of student learning).

By far the most challenging condition researchers must meet in order to make a causal inference is eliminating other plausible alternative causes. Consider a study in which the effect of two different teaching approaches (active and passive) is assessed. Suppose the researcher assigns students to teaching conditions by having all men in one group and all women in the other. If this were done, any difference between the two groups could be due either to the teaching method *or* to the gender of the students. Thus, the researcher would not be able to determine whether the difference in performance between the two groups was due to the independent variable (active or passive learning) or to the alternative explanation of students' gender. Said more formally, the independent variable of teaching method would be "confounded" with the independent variable of gender. **Confounding** occurs when two potentially effective independent 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.

Key Concept

Researchers seek to explain the causes of phenomena by conducting experiments. However, even when a carefully controlled experiment allows the researcher to form a causal inference, additional questions remain. One important question concerns the extent to which the findings of the experiment apply only to the people who participated in the experiment. Researchers often seek to generalize their findings to describe people who did not participate in the experiment.

Most of the participants in psychology research are introductory psychology students in colleges and universities. Are psychologists developing principles that apply only to college freshmen and sophomores? Similarly, laboratory research is often conducted under more controlled conditions than are found in natural settings. Thus, an important task of the scientist is to determine whether laboratory findings generalize to the "real world." Some people automatically assume that laboratory research is useless or irrelevant to real-world concerns. However, as we explore research methods throughout this text, we will see that these views about the relationship between laboratory science and the real world are not helpful or satisfying. Instead, psychologists recognize the importance of both: Findings from laboratory experiments help to explain phenomena, and this knowledge is applied to real-world problems in research and interventions.

Application

- In applied research, psychologists apply their knowledge and research methods to improve people's lives.
- Psychologists conduct basic research to gain knowledge about behavior and mental processes and to test theories.

The fourth goal of research in psychology is application. Although psychologists are interested in describing, predicting, and explaining behavior and mental processes, this knowledge doesn't exist in a vacuum. Instead, this knowledge exists in a world in which people suffer from mental disorders and are victims of violence and aggression, and in which stereotypes and prejudices impact how people live and function in society (to name but a few problems we face). The list of problems in our world may at times seem endless, but this shouldn't discourage us. The breadth of psychologists' research questions and findings provides many ways for researchers to help address important aspects of our lives and to create change in individuals' lives.

Key Concept

Research on creating change is often called "applied research." In **applied research,** psychologists conduct research in order to change people's lives for the better. For people suffering from mental disorders, this change may occur through research on therapeutic techniques. However, applied psychologists are involved with many different types of interventions, including those aimed at improving the lives of students in schools, employees at work, and individuals in the community. On the other hand, researchers who conduct **basic research** seek primarily to understand behavior and mental processes. People often describe basic research as "seeking knowledge for its own sake." Basic research is typically carried out in a laboratory setting with the goal of testing a theory about a phenomenon.

Key Concept

Throughout the history of psychology, tension has existed between basic research and applied research. Within the past several decades, however, researchers have increased their focus on important, creative applications of psychological principles for improving human life (Zimbardo, 2004). In fact, the application of well-known principles of psychology—discovered through basic research—is now so pervasive that people tend to forget the years of basic research in laboratories that preceded what we now understand to be commonplace. For example, the use of positive reinforcement techniques, psychological testing and therapies, and self-help practices has become part of everyday life. In addition, the application of psychological principles is becoming increasingly important in education, health, and criminal justice settings. To see some of the many applications of psychology in our everyday life, check out this website: www.psychologymatters.org.

One important factor ties together basic and applied research: the use of theories to guide research and application in the real world. In the next section we describe how psychological theories are developed.

Scientific Theory Construction and Testing

- Theories are proposed explanations for the causes of phenomena, and they vary in scope and 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.
- Intervening variables are concepts used in theories to explain why independent and dependent variables are related.
- Successful scientific theories organize empirical knowledge, guide research by offering testable hypotheses, and survive rigorous testing.
- 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 theories about the nature of behavior and mental processes, as well as about the reasons people and animals behave and think the way they do. A psychological theory can be developed on different levels; for example, the theory can be developed on either a physiological or a conceptual level (see Anderson, 1990; Simon, 1992). A physiologically based theory of schizophrenia would propose biological causes such as specific genetic carriers. A theory developed on a conceptual level would more likely propose psychological causes such as patterns of emotional conflict or stress. It would also be possible for a theory of schizophrenia to include both biological and psychological causes. The propositions contained in theories may be expressed as verbal statements, as mathematical equations, or as computer programs.

Theories often differ in their scope—the range of phenomena they seek to explain. Some theories attempt to explain specific phenomena. For example, Brown and Kulik's (1977) theory attempted to explain the phenomenon of "flashbulb memory." A flashbulb memory refers to the finding that we remember very specific personal circumstances surrounding particularly surprising and emotional events in our lives, such as the horrific events of September 11, 2001. Other theories have much broader scope as they try to describe and explain more complex phenomena such as love (Sternberg, 1986) or human cognition (Anderson, 1990, 1993; Anderson & Milson, 1989). In general, the greater the scope of a theory, the more complex it is likely to be. Most theories in contemporary psychology tend to be relatively modest in scope, attempting to account only for a limited range of phenomena.

Scientists develop theories from a mixture of intuition, personal observation, and known facts and ideas. The famous philosopher of science Karl Popper (1976) suggested that truly creative theories spring from a combination of intense interest in a problem and critical imagination—the ability to think critically and "outside the box." Researchers begin constructing a theory by considering what is known about a problem or research question and also looking for errors or what is missing. The approach is similar to the one we described in Chapter 1 for getting started in research and forming hypotheses.

Key Concept

Although theories differ in their level of explanation and scope, amid these differences there are commonalities that define all theories. 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*. For example, a theory of flashbulb memory needs to state exactly what a flashbulb memory is, including a description of how a flashbulb memory differs from typical memories. The theory would also need to include descriptions of relationships, such as the relationship between degree of emotional involvement and amount remembered. Finally, the theory would also have to 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). Such was the case in Talarico and Rubin's (2003) findings for students' memories of the September 11, 2001, terrorist attacks; despite a decrease in the accuracy of their memories over time, participants maintained confidence in their very vivid memories.

The major functions of a theory are to *organize* empirical knowledge and to *guide* research (Marx, 1963). Even in relatively specific areas of research such as flashbulb memories, many studies have been done. As the scope of a research area increases, so does the number of relevant studies. Scientific theories are important because they provide a logical organization of many research findings and identify relationships among findings. This logical organization of findings guides researchers as they identify testable hypotheses for their future research.

Theories frequently require that we propose intervening processes to account for observed behavior (Underwood & Shaughnessy, 1975). These intervening processes provide a link between the independent variables researchers manipulate and the dependent variables they subsequently measure. Because these processes "go between" the independent and dependent variables, they are called *intervening variables*. You probably are familiar with what we mean by an intervening variable if you think about your computer use. As you press keys on the keyboard or click the "mouse," you see (and hear) various outcomes on the monitor, printer, and from the speakers. Yet it isn't your keystrokes and mouse clicks that *directly* cause these outcomes; the intervening variable is the "invisible" software that serves as a connection between your keystrokes and the outcome on your monitor.

Intervening variables are like computer software. Corresponding to the connection between keystrokes and what you see on your monitor, intervening variables connect independent and dependent variables. Another familiar example from psychology is the construct of "thirst." For example, a researcher might manipulate the number of hours participants are deprived of liquid and, after the specified time, measure the amount of liquid consumed. Between the deprivation time and the time participants are allowed to drink liquid, we may say that the participants are "thirsty"—the psychological experience of needing to replenish body fluids. Thirst is a construct that allows theorists to connect variables such as the number of hours deprived of liquid (the independent variable) and the amount of liquid consumed (the dependent variable). *Intervening variables such as thirst not only link independent and dependent variables; intervening variables also are used to explain why the variables are connected.* Thus, intervening variables play an important role when researchers use theories to explain their findings.

Intervening variables and theories are useful because they allow researchers to identify relationships among seemingly dissimilar variables. Other independent variables likely influence "thirst." Consider, for example, a different independent variable: amount of salt consumed. On the surface, these two independent variables—number of hours deprived of liquid and amount of salt consumed—are very dissimilar. However, both influence subsequent consumption of liquid and can be explained by the intervening variable of thirst. Other independent variables related to liquid consumption include amount of exercise and temperature; the more exercise or the higher the temperature, the more people are "thirsty" and the more liquid they consume. Although these examples emphasize independent variables, it's important to note that dependent variables also play a role in theory development. Thus, rather than measuring "liquid consumption" as the dependent variable, inventive researchers may measure other effects related to the psychological experience of thirst. For example, when deprived of liquid, individuals may go to greater efforts to obtain liquid or may even drink liquids that taste bitter. Thus, effort to obtain liquids or the amount of bitterness in the liquid could be measured as dependent variables.

Intervening variables are critical to theory development in psychology. In our example, the apparently dissimilar variables of liquid deprivation, salt consumption, exercise, temperature, liquid consumption, effort to obtain liquid, and taste of liquids can be united in one theory that relies on the intervening variable "thirst." Other examples of intervening variables—and theories—abound in psychology. The intervening variable "depression," for example, connects the factors theorized to cause depression (e.g., neurological factors, exposure to trauma) and the various symptoms (e.g., sadness, hopelessness, sleep and appetite disturbance). Similarly, "memory" as an intervening variable is used to explain the relationship between the amount (or quality) of time spent studying and later performance on a test. As you will learn in your study of psychology, intervening variables provide the key that unlocks the complex relationships among variables.

How we evaluate and test scientific theories is one of the most difficult issues in psychology and philosophy (e.g., Meehl, 1978, 1990a, 1990b; Popper, 1959). Kimble (1989) has suggested a simple and straightforward approach. He says, "The best theory is the one that survives the fires of logical and empirical testing" (p. 498). Scientists first evaluate a theory by considering whether it is logical. That is, they determine whether the theory makes sense and whether its propositions are free of contradictions. The logical consistency of theories is tested through the lens of the critical eye of the scientific community.

The second "fire" that Kimble (1989) recommends for evaluating theories is to subject hypotheses derived from a theory to empirical tests. 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, to confirming or disconfirming scientific theories. For example, a theory, especially a complex one, may produce many specific testable hypotheses. A theory is not likely to fail on the basis of a single test (e.g., Lakatos, 1978). Moreover, theories may include concepts that are not adequately defined or suggest complex relationships among intervening variables and behavior. Such theories may have a long life, but their value to science is questionable (Meehl, 1978). Ultimately, the scientific community determines whether any test of a theory is definitive.

In general, theories that provide *precision of prediction* are likely to be much more useful (Meehl, 1990a). For example, a theory that predicts that children will typically demonstrate abstract reasoning by age 12 is more precise (and testable) in its predictions than a theory that predicts the development of abstract reasoning by ages 12 to 20. When constructing and evaluating a theory, scientists also place a premium on parsimony (Marx, 1963). The *rule of parsimony* is followed when the simplest of alternative explanations is accepted. Scientists prefer theories that provide the simplest explanations for phenomena.

In summary, a good scientific theory is one that is able to pass the most rigorous tests. Somewhat counterintuitively, rigorous testing will be more informative when researchers do tests that seek to *falsify* a theory's propositions than when they do tests that seek to confirm them (Shadish, Cook, & Campbell, 2002). Although tests that confirm a particular theory's propositions do provide support for the specific theory that is being tested, confirmation logically does not rule out other, alternative theories of the same phenomenon. Tests of falsification are the best way to prune a theory of its dead branches. Constructing and evaluating scientific theories is at the core of the scientific enterprise and is absolutely essential for the healthy growth of the science of psychology.

SUMMARY

As an approach to knowledge, the scientific method is characterized by a reliance on empirical procedures, rather than relying only on intuition, and by an attempt to control the investigation of those factors believed responsible for a phenomenon. Scientists gain the greatest control when they conduct an experiment. In an

experiment, those factors that are systematically manipulated 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 variables are called dependent variables.

Scientists seek to report results in an unbiased and objective manner. This goal is enhanced by giving operational definitions to concepts. Psychological researchers refer to concepts as "constructs." Scientists also use instruments that are as accurate and precise as possible. Phenomena are quantified with both physical and psychological measurement. Scientists seek measures that have both validity and reliability. 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.

The goals of the scientific method are description, prediction, explanation, and application. 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 and explanation are 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. Even when a carefully controlled experiment allows the researcher to form a causal inference, additional questions remain concerning the extent to which the findings may generalize to describe other people and settings. In applied research, psychologists strive to apply their knowledge and research methods to improve people's lives. Basic research is conducted to gain knowledge about behavior and mental processes and to test theories.

Scientific theory construction and testing are at the core of the scientific approach to psychology. A theory is defined as a logically organized set of propositions that serves to define events, describe relationships among these events, and explain the occurrence of the events. Theories have the important functions of organizing empirical knowledge and guiding research by offering testable hypotheses. Intervening variables are critical to theory development in psychology because these constructs allow researchers to explain the relationships between independent and dependent variables.

Key Concepts

control 30
experiment 32
independent variable 32
dependent variable 33
construct 36
operational definition 36
validity 39
reliability 39
correlation 45
causal inference 47
confounding 48
applied research 49
basic research 49
theory 51

Review Questions

1 For each of the following characteristics, distinguish between the scientific approach and everyday approaches to knowledge: general approach, observation, reporting, concepts, instruments, measurement, and hypotheses.
2 Differentiate between an independent variable and a dependent variable, and provide an example of each that could be used in an experiment.
3 What is the major advantage of using operational definitions in psychology? In what two ways has the use of operational definitions been criticized?
4 Distinguish between the accuracy and the precision of a measuring instrument.
5 What is the difference between the validity of a measure and the reliability of a measure?
6 Which three types of hypotheses lack the critical characteristic of being testable?
7 Identify the four goals of the scientific method and briefly describe what each goal is intended to accomplish.
8 Distinguish between the nomothetic approach and the idiographic approach in terms of who is studied and the nature of the generalizations that are sought.
9 Identify two differences between quantitative and qualitative research.
10 What are researchers able to do when they know that two variables are correlated?
11 Give an example from a research study described in the text that illustrates each of the three conditions for a causal inference. [You may use the same example for more than one condition.]
12 What is the difference between basic and applied research?
13 What is an intervening variable? Propose a psychological construct that could serve as an intervening variable between "insult" (present/absent) and "aggressive responses." Explain how these variables might be related by proposing a hypothesis that includes your intervening variable.
14 Describe the roles of logical consistency and empirical testing in evaluating a scientific theory.
15 Explain why rigorous tests of a theory that seek to falsify a theory's propositions can be more informative than tests that seek to confirm a theory's propositions.

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 that occur as children gain experience with large playground equipment (e.g., slides, swings, climbing walls). For a span of 8 weeks, preschool children were assigned to 4, 6, or 8 hours per week for time allowed on the equipment. She then tested their motor coordination by asking them to skip, jump, and stand on one foot.

C A developmental psychologist was interested in the amount of verbal behavior very young children displayed depending on who else was present. The children in the study were 3 years old. These children were observed in a laboratory setting for a 30-minute period. Half of the children 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

(*continued*)

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 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. (For the sake of this hypothetical problem, we are ignoring the fact 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 (average) 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.

A Identify the independent variable of interest (and its levels) and a potentially relevant independent variable with which the primary independent variable is confounded. Explain clearly how the confounding occurred.

B State exactly what conclusion about the effect of the drug Speedo can be supported by the evidence presented.

C Finally, suggest ways in which the study could be done so that you could make a clear conclusion about the effectiveness of the drug Speedo.

3 The *New York Times* reported the results of a 2-year, $1.5 million study by researchers at Carnegie Mellon University funded by the National Science Foundation and major technology companies. There were 169 participants in the study drawn from the Pittsburgh area. The researchers examined the relationship between Internet use and psychological well-being. A director of the study stated that the study did not involve testing extreme amounts of Internet use. The participants were normal adults and their families. On average, for those who used the Internet the most, psychological well-being was the worst. For example, 1 hour a week of Internet use led to slight increases on a depression scale and on a loneliness scale and a reported decline in personal interaction with family members. The researchers concluded that Internet use appears to cause a decline in psychological well-being. They suggested that users of the Internet were building shallow relationships that led to an overall decline in feelings of connection to other people.

A The researchers claim that use of the Internet leads to a decline in people's well-being. What evidence is present in this summary of the report to meet the conditions necessary for drawing this causal inference and what evidence is lacking?

B What sources beyond this question would you want to check before reaching a conclusion about the findings reported here? [You might begin with the *New York Times* piece "The Lonely Net," August 30, 1998, and the *Washington Post* piece "Net Depression Study Criticized," September 7, 1998.]

C What effect would there be on your evaluation of this study given that it was funded in part by major technology companies?

4 A study was done to determine whether taking notes in a developmental psychology course affected students' test performance. Students recorded their notes over the entire semester in a 125-page study guide. The study guide included questions on course content covered both in the textbook and in class lectures. Students' notes were measured using three dimensions: completeness, length, and accuracy. Results of the study indicated that students with more accurate notes performed better on essay and multiple-choice tests in the course than did students with less accurate notes. Based on these findings, the researchers suggested that instructors should use instructional techniques such as pausing for brief periods during the lecture and asking questions to clarify information. The researchers argued that these techniques could facilitate the accuracy of the notes students take in class, and that accurate note taking could contribute significantly to students' overall success in college courses.

A What evidence is present in this report to meet the conditions for a causal inference between accuracy of students' notes and their test performance? What evidence is lacking? (Be sure to identify clearly the three conditions for a causal inference.)

B Identify a goal of the scientific method that could be met on the basis of findings of this study.

Answer to Stretching Exercise

1 The independent variable in this study is the emotion condition participants experienced after completing the hand-eye coordination task. There were three levels: gratitude, amusement, and neutral. The dependent variable was the number of minutes participants helped by completing the confederate's survey.

2 An alternative explanation for the study's finding is that participants simply felt good when the confederate fixed the computer problem and therefore helped more at the end of the experiment. To show that the specific emotion of gratitude was important, the researchers used one experimental condition, the amusing video condition, to control for positive emotions in general. That is, if simply positive emotions cause greater helping, then these participants should show greater helping also. Because only participants in the gratitude condition showed the greatest helping, the researchers can argue that gratitude specifically cause increased helping.

Answer to Challenge Question 1

A Independent variable (IV): hours of food deprivation with four levels (0, 8, 16, 24); dependent variable (DV): time (e.g., in minutes) animals spent in activity wheel

B IV: time on playground equipment with three levels: 4, 6, or 8 hours per week; DV: scores on test of motor coordination

C IV: additional person present with two levels (adult, child); DV: number, duration, and complexity of child's verbal utterances