# 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. One of the best ways to understand the scientific method as a means of gaining knowledge 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 and are summarized in the following pages. Collectively, the characteristics listed under "Scientific" define the scientific method.

**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 |
| Concepts: | Ambiguous, with surplus meanings | Clear definitions, operational specificity |
| Reporting: | Biased, subjective | Unbiased, objective |
| 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).

## General Approach and Attitude

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 using intuition. This usually means that we act on the basis of what "feels right" or what "seems reasonable." Although intuition can be valuable when we have little other information, intuition is not always correct.

When we rely on intuition to make judgments we often fail to recognize that our perceptions may be distorted by cognitive biases, or that we may not have considered all available evidence (see Box 2.1; Kahneman & Tversky, 1973; Tversky & Kahneman, 1974). By using the scientific method, psychologists seek to avoid the *confirmation bias*—our natural tendency to seek evidence that is consistent with our intuitions and ignore or deny contradictory evidence. Confirmation bias drives people's choice of news programs (e.g., Fox, CNN, MSNBC) and motivates people to avoid information that challenges their preexisting attitudes, beliefs, and behaviors, even when avoiding information causes them to be wrong (Hart et al., 2009).

Psychological research has demonstrated many examples of how confirmation bias occurs. For example, even young children are susceptible to the cognitive error called *illusory correlation,* which is a tendency to perceive relationships between events when none exists. In one study (Susskind, 2003), researchers showed children many pictures of men and women performing stereotypical (e.g., a woman knitting), counterstereotypical (e.g., a man knitting), and neutral behaviors (e.g., reading a book). When asked to estimate how many of each type

**BOX 2.1**

### WE'RE NOT AS SMART (OR GOOD) AS WE THINK

Psychologists have long recognized that human thinking is prone to errors and biases. These biases affect the way we make predictions and decisions, and how we interact with people. A number of books targeted for a general audience describe this research, with several of them reaching bestseller lists. Check out these books:

- *Blind Spot: Hidden Biases of Good People* by Mahzarin R. Banaji and Anthony G. Greenwald (Delacorte, 2013)
- *The Invisible Gorilla: How Our Intuitions Deceive Us* by Christopher Chabris and Daniel Simons (Crown, 2010)
- *The Signal and the Noise: Why So Many Predictions Fail—But Some Don't* by Nate Silver (The Penguin Press, 2012)
- *Thinking, Fast and Slow* by Daniel Kahneman (Farrar, Straus, and Giroux, 2011)
- *You Are Not So Smart: Why Your Memory Is Mostly Fiction, Why You Have Too Many Friends on Facebook, and 46 Other Ways You're Deluding Yourself* by David McRaney (Gotham, 2011)

of picture they saw, children overestimated the number of stereotypical pictures. By noticing pictures consistent with their beliefs more than contradictory pictures, they confirmed their stereotypes about men and women's behavior.

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. Research at first may be guided by the scientist's intuition. Eventually, however, the scientist strives to be guided by the empirical evidence that direct observation and experimentation provide.

## Observation

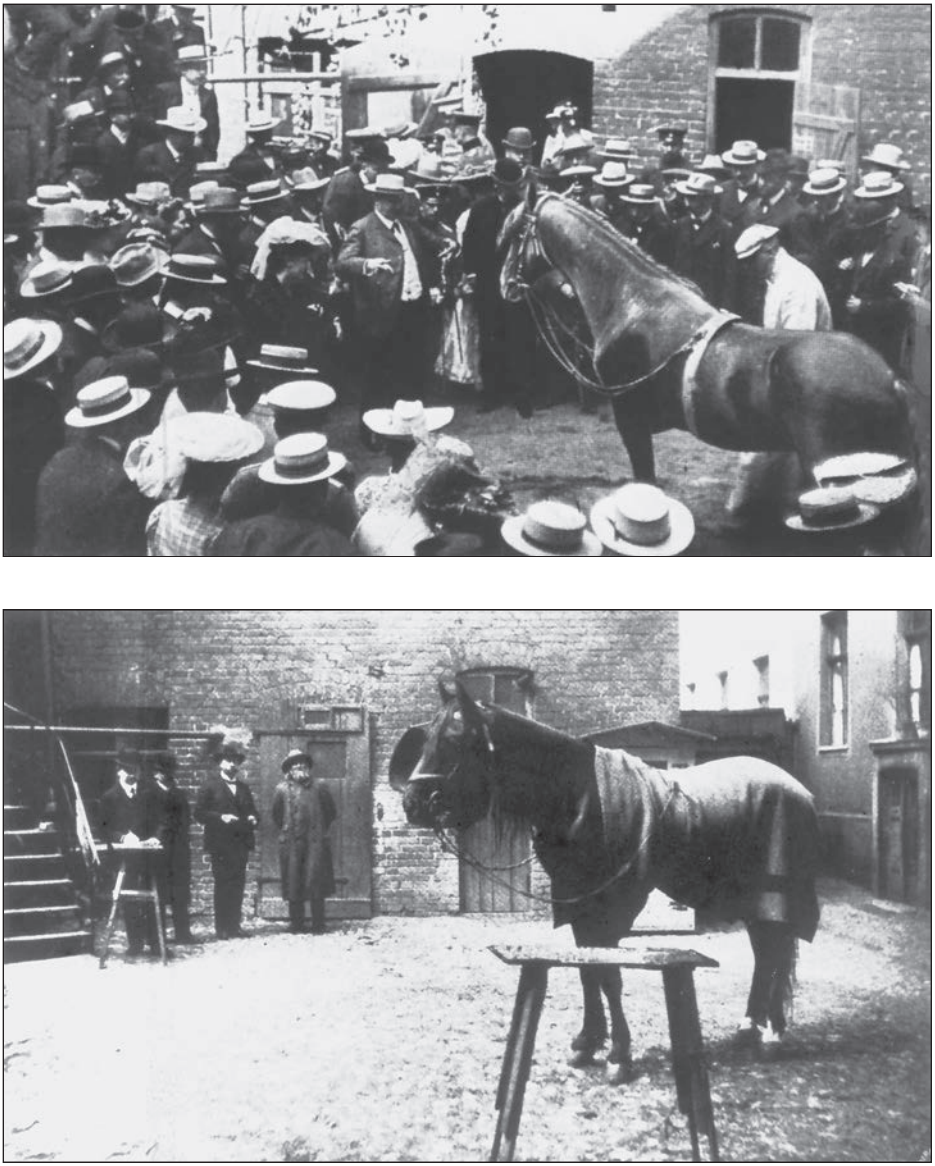
We can learn a great deal about behavior by simply observing the actions of others. However, everyday observations are not always made carefully or systematically. Most people do not attempt to control or eliminate factors that might influence the events they are observing. As a result, we often make 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 controlled observation, scientists gain a clearer picture of the factors that produce

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



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 2.2 describes an example of how the story of Clever Hans from over 100 years ago informs scientists even today.

*Key Concept*

Scientists gain the greatest control when they conduct an experiment. In an **experiment,** scientists manipulate one or more factors and observe the effects of

BOX 2.2

## 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. More recent studies suggest mixed results (e.g., Gordon et al., 2008; McCulloch et al., 2006). 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 Concept*

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

[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 6, there are important differences between manipulated and selected independent variables.

effect of drinking alcohol on the ability to process complex information quickly 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 behavior (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.

## 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 clearly define our concepts.

In everyday conversation we often get by without worrying too much 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. 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, and people will provide different definitions. Clearly, to answer the question of whether intelligence is mostly inherited or learned, we need to have an exact definition that all parties can accept.

*Key Concept*

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

## 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, another 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?

*Key Concepts*

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 communicating about constructs has not escaped criticism. One problem is that if we don't like one operational definition of intelligence, then simply give intelligence another operational definition. Does this mean that there are as many kinds of intelligence as there are operational definitions? The answer, unfortunately, is that we don't really know. To determine whether a different procedure or test 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. This is particularly relevant in cross-cultural research where, for example, a paper-and-pencil test of intelligence may tap into knowledge that is specific to a particular cultural context. 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 one test compare to performance on other tasks that are commonly accepted as measures 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.

## 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.

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

**FIGURE 2.2** How would you describe this scene?

to someone what you see there? One way to describe this scene is to say that two people are running along a path. You might also describe this scene as two 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 two 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 2.2 as running could be considered an inference, the actual observation being that two 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.

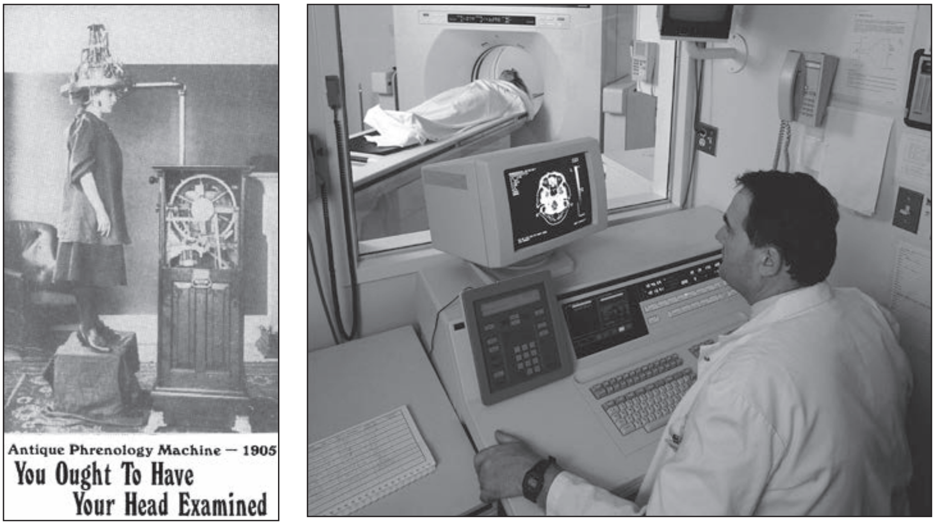
Scientific reporting seeks to be *unbiased* and *objective*. One way to determine whether a report is unbiased is to see if it can be verified by an independent observer. This is called "interobserver agreement" (see Chapter 4). Unfortunately, many biases are subtle and not always detected even in scientific 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.

## Instruments

You depend on instruments to measure events more than you probably realize. For example, you rely on the speedometer in a car and the clock in your bedroom, and you can appreciate the problems that arise when these instruments are inaccurate. *Accuracy* refers to the difference between what an instrument says is true and what is known to be true. Inaccurate clocks can make us late, and inaccurate speedometers can earn us traffic tickets. The accuracy of an instrument is determined by *calibrating* it, or checking it with another instrument known to be true. In addition, 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.

We also need instruments to measure behavior. You can be assured that the precision, and even the accuracy, of instruments used in psychology have improved significantly since 1879, the founding of the first psychology laboratory. Today, many sophisticated instruments are used in contemporary psychology (Figure 2.3). Psychophysiology experiments (e.g., when assessing a person's arousal level) require instruments that give accurate measures of such internal states as heart rate and blood pressure. Questionnaires and tests are popular instruments used by psychologists to measure behavior. So, too, are

**FIGURE 2.3** Scientific instruments used in psychology have improved dramatically in their precision and accuracy.




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

Scientists use two types of measurements to record the 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 measurement instrument. For example, length is a dimension with agreed-upon standards for units of length (e.g., inches, meters) and measurement instruments (e.g., a ruler). Similarly, units of weight and time represent physical measurement.

Most psychological research, however, does not involve physical measurement. Rulers do not exist for measuring psychological constructs such as beauty, aggression, or intelligence. These dimensions require 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*. For example, if several independent observers agree that a certain action warrants a rating of 3 on a 7-point rating scale of aggression, that is a psychological measurement of the aggressiveness of the action.

*Key Concept*

Measurements must be 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. Suppose a researcher defines intelligence in terms of how long a person can balance a ball on his or her 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 valid measure of intelligence. The validity of a measure is supported when people do as well on it as on other tasks 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 other accepted measures of intelligence.

*Key Concept*

The **reliability** of a measurement is indicated by its consistency. There are several kinds of reliability. When we speak of instrument reliability, we are discussing whether an instrument works consistently. A car that sometimes starts and sometimes doesn't is not very reliable. Observations made by two or more independent observers are said to be reliable if they agree—that is, if the observations are consistent from one observer to another. For example, researchers asked 50 college students to examine photos of students from a different university and rate the extent to which those students appeared trustworthy using a 7-point scale (1 = *not at all* and 7 = *very trustworthy*). Results indicated very high agreement among the raters, indicating that trustworthiness can be reliably measured (Rule, Krendl, Ivcevic, & Ambady, 2013). Interestingly, the students who had been photographed also completed a set of research tasks, including a brief test in which they were given an opportunity to cheat by using more time than allowed. More than half (57%) of the students cheated on the test. The researchers found that ratings of trustworthiness were completely unrelated to whether students cheated or not. Thus, reliable measurement does not always mean the measure is valid.

The validity and reliability of measurements are central issues in psychological research. We will describe various ways in which researchers determine the reliability and validity of their measures 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, an emerging area of psychological research asks, Why do people purchase "green" products, especially when these products are often more expensive and may be less luxurious or effective than conventional, nongreen products? An example is the successful Toyota Prius, which is as expensive as cars that are more comfortable and perform better. One *hypothesis* for green purchases relates to altruism, the tendency toward selfless acts that benefit others (Griskevicius, Tybur, & Van den Bergh, 2010). Purchasing green products can be seen as altruistic because the environment and society benefit, with a greater cost to the selfless purchaser.

Recent theorists describe "competitive altruism," in which individuals are altruistic because being seen as prosocial and selfless enhances one's reputation and status in society (Griskevicius et al., 2010). Thus, altruistic acts, such as purchasing green products, may signal one's higher status—that one has the time, energy, wealth, and other resources to behave altruistically. Griskevicius et al. hypothesized that activating (i.e., making prominent) people's desire for status should lead them to choose green products over more luxurious nongreen products.

Griskevicius et al. (2010) conducted three experiments to test their hypothesis. In each, they manipulated college student participants' motivation for status using two conditions: status and control. Status motives were activated by having participants in this condition read a short story about graduating from college, searching for a job, and then working for a desirable company with opportunities for promotion. In the control condition, participants read a story about searching for a lost concert ticket, finding it, and then attending the concert. After reading the story, participants believed they were completing a second, unrelated study about consumer preferences. They identified items they would likely purchase (e.g., car, dishwasher, backpack); in each case, a green product was paired with a nongreen, more luxurious item. Griskevicius et al. found that compared to the control condition, activating status motives increased the likelihood that participants would choose green products over the nongreen products (Experiment 1). Furthermore, the preference for green products occurred only when status-motivated participants imagined shopping in public, but not in private (online) situations (Experiment 2), and when green products cost more than nongreen products (Experiment 3).

At a theoretical level, a hypothesis may offer a reason (the "why") for the way particular variables are related. Griskevicius and his colleagues found a relationship between two variables: status motives and the likelihood of purchasing green products. Based on theories of competitive altruism, these variables are related because people gain social status when they are seen to behave altruistically, such as when purchasing green products. One practical implication for this finding is that sales of green products may be enhanced by linking these products with high status (e.g., celebrity endorsements), rather than by emphasizing the plight of the environment or by making green products less expensive.

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 or measured.* For example, to say that a would-be assassin shot a prominent figure or celebrity because the assassin is mentally disturbed is not a testable hypothesis unless we can agree on a definition of "mentally disturbed." 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. You may have learned in a psychology course that many of Freud's hypotheses are not testable. This is because there are no clear operational definitions and measures for key constructs in Freud's theories, such as *id*, *ego,* and *superego.*

*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 controlled by 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 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 (see Table 2.2).

### 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* (5th ed., 2013), also known as DSM-5 (see Figure 2.4). Consider, as one example, the criteria used to define the disorder labeled selective mutism.

**TABLE 2.2** FOUR GOALS OF PSYCHOLOGICAL RESEARCH

| Goal | What Is Accomplished | Example |
|---|---|---|
| **Description** | Researchers define, classify, catalogue, or categorize events and relationships to describe mental processes and behavior. | Psychologists describe symptoms of helplessness in depression, such as failure to initiate activities and pessimism regarding the future. |
| **Prediction** | When researchers identify correlations among variables they are able to predict mental processes and behavior. | As level of depression increases, individuals exhibit more symptoms of helplessness. |
| **Explanation** | Researchers understand a phenomenon when they can identify the cause(s). | Participants exposed to unsolvable problems become more pessimistic and less willing to do new tasks (i.e., become helpless) than participants who are asked to do solvable problems. |
| **Application** | Psychologists apply their knowledge and research methods to change people's lives for the better. | Treatment that encourages depressed individuals to attempt tasks that can be mastered or easily achieved decreases depressives' helplessness and pessimism. |

Based on Table 1.2, Zechmeister, Zechmeister, & Shaughnessy, 2001, p. 12.

**Diagnostic Criteria for Selective Mutism**

**(a)** Consistent failure to speak in specific social situations in which there is an expectation for speaking (e.g., at school) despite speaking in other situations.
**(b)** The disturbance interferes with educational or occupational achievement or with social communication.
**(c)** The duration of the disturbance is at least one month (not limited to the first month of school).
**(d)** The failure to speak is not attributable to a lack of knowledge of, or comfort with, the spoken language required in the social situation.
**(e)** The disturbance is not better explained by a communication disorder (e.g., childhood-onset fluency disorder) and does not occur exclusively during the course of autism spectrum disorder, schizophrenia, or another psychotic disorder. (DSM-5, 2013, p. 195)

The diagnostic criteria used to define selective mutism provide an operational definition for this disorder. Selective mutism is relatively rare; thus, we typically rely on "case studies" to learn about individuals with this disorder. 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. For example, research suggests that selective mutism is seen most often among young children, and because of the strong association between selective mutism and social anxiety,

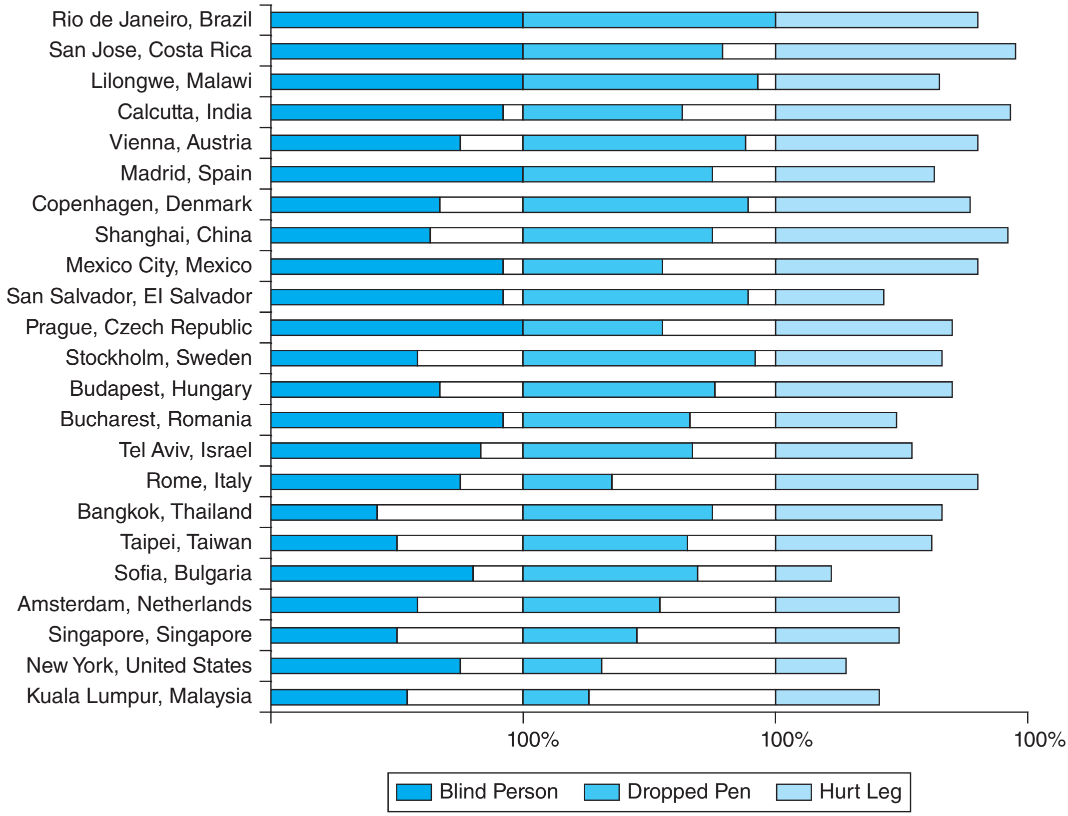
**FIGURE 2.4** Clinicians classify mental disorders according to the criteria found in the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders.*

the diagnosis is now classified among anxiety disorders (DSM-5, 2013). In Chapter 9 we will consider a research design used by psychologists to treat a young girl with selective mutism.

Science, including psychological science, develops descriptions of phenomena using the *nomothetic approach*. "Nomothetic" refers to the discovery of general scientific laws. Psychologists try to establish broad generalizations and general laws of behavior 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, one goal of a large cross-cultural study was to describe the extent of helping in large cities around the world (Levine, Norenzayan, & Philbrick, 2001). Researchers created three helping situations in downtown settings to record whether citizens helped. In each country, an experimenter (1) acted the role of a blind person attempting to cross a street; (2) failed to notice he had accidentally dropped a pen while walking; and (3) while wearing a large leg brace, accidentally dropped and struggled to pick up a pile of magazines. In all, 1,198 people were observed in one of these situations across 23 cities. Results for the study are shown in Figure 2.5. On average, people in Rio de Janeiro demonstrated the greatest amount of helping and citizens of Kuala Lumpur the least. New York City ranked second to last. These findings do not indicate that *all* people in Rio de Janeiro are helpful, nor are all people in Kuala Lumpur

**FIGURE 2.5** Percentage of people who helped in three situations (assisting a blind person, picking up a dropped pen, helping someone with hurt leg) in 23 large cities around the world, rank-ordered according to overall amount of helping, with Rio de Janeiro the highest and Kuala Lumpur the lowest. (From Levine et al., 2001, Table 2.)




unhelpful. These findings indicate that *in general* or *on average*, the extent to which individuals help differs across the 23 cities in this study.

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 single-case research, which we will describe in Chapter 9.

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 attempting to answer research questions with a large group of research participants. 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 described using 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, anthropologists, and psychologists. While there is no single way to conduct qualitative research studies (Silverman, 2011), researchers using this approach typically look for meaningful themes and categories in a narrative record, then provide a verbal summary of their observations. Qualitative research often focuses on events in their context and frequently is based on personal interviews and comprehensive records obtained from direct observation of behavior. More recent analyses include Internet sources, such as messages in chat rooms. When conducting interviews, qualitative researchers may ask participants to describe their experiences in ways that are meaningful to *them*, rather than asking participants to answer questions using 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 9 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. Many important questions in psychology 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 the answer to all of these questions is "yes." 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. In addition, an important occupation for many psychologists involves psychological testing, in which tests are used to predict future behavior and performance (e.g., on the job, in specific vocations). You are already familiar with some of these, such as the tests used for admission into college and professional schools.

*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. Researchers asked how well ratings of teachers by students *not* enrolled in the class would correlate with end-of-the-semester evaluations made by students *in* the class (Ambady & Rosenthal, 1993). They showed video clips (without sound) of teachers to a group of female undergraduates 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 students' end-of-the-semester teacher evaluations. 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. Thus, we can predict course evaluations of teachers' affective behavior (e.g., likableness) based on ratings of briefly depicted videotaped behavior. These results indicate that observers can make relatively accurate judgments of affective behavior (e.g., likeableness) very quickly.

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 some people 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. Furthermore, we would never argue that an animal's strange behavior caused an earthquake.

In their study of cross-cultural differences in helping behavior, Levine et al. (2001) noted that the extent of helping in a city could be predicted by an indicator of economic well-being concerning the purchasing power of citizens' income ("how far a dollar will go"). People were more likely to help when the purchasing power in a city was *lower*. Although they cannot argue that low purchasing power *causes* helping, these researchers speculated that when a country does not have a strong economy, traditional value systems may mandate helping behaviors.

Building on this idea that traditional values may explain differences in helping across cultures, another set of researchers investigated the concept of "embedded cultures" (Knafo, Schwartz, & Levine, 2009). An embedded culture is one in which individuals gain meaning in life through their identification with their family or an in-group. Embedded individuals focus on the welfare of group members, with less concern for people outside the group (such as strangers with a hurt leg or needing help locating a pen). These investigators measured cultural embeddedness for the 23 cities for which helping behaviors were observed in the earlier study. As hypothesized, the more embedded a culture, the less likely were people to help the stranger. Although this correlational finding supports the proposed explanation that traditional, embedded values influence people's helping behavior, additional evidence is needed to make a statement that these values *cause* differences in helping across cultures.

## 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. Many people use the word "experiment" when speaking about research in general, but this is incorrect. 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 6, 7, and 8 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 use stern discipline and 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 we assume parents 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 young children vary in how active and aggressive they are and that the child'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.

A key component for making a causal inference is comparing outcomes in two or more conditions. For example, suppose a teacher wished to demonstrate that using personal response systems ("clickers") in the classroom helps students learn. She could ask questions in class that require a response on the clicker, and then describe the test performance of students following this classroom technique. 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 clickers *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 clickers 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 (clicker and no-clicker) is assessed. Suppose the researcher assigns students to teaching conditions by using clickers in the morning section of a course, and does not use clickers in the afternoon section. If this were done, any difference between the two groups could be due either to the teaching method *or* to the different sections, morning *vs.* afternoon. Thus, the researcher would not be able to determine whether the difference in performance between the two groups was due to the independent variable she tested (clicker or no-clicker) or to the alternative explanation of differences among students in the two sections. Said more formally, the independent variable of teaching method would be "confounded" with the independent variable of course section. **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 want to generalize their findings to describe people who did not participate in the experiment.

Many 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.apa.org/research/action.

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 using different levels of explanation; 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.

Theories differ in their scope—the range of phenomena they seek to explain. Theories with a broad scope try to describe and explain complex phenomena such as love or human cognition. In general, the greater the scope of a theory, the more complex it is likely to be. Most theories in contemporary psychology, however, tend to be relatively modest in scope, attempting to account for a limited range of phenomena. For example, the theory of "flashbulb memory" attempts to explain people's memory for very specific personal details surrounding surprising and emotional events, such as the horrific events of September 11, 2001 (Brown & Kulik, 1977).

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 (see Table 2.3). We can offer the following formal definition of a scientific **theory:** *a logically*

**TABLE 2.3** CHARACTERISTICS OF THEORIES

| | |
|---|---|
| **Definition** | A theory is a logically organized set of propositions that serves to define events, describe relationships among these events, and explain the occurrence of these events. |
| **Scope** | Theories differ in the breadth of events they seek to explain, from very specific phenomena (e.g., flashbulb memory) to complex phenomena (e.g., love). |
| **Functions** | A theory organizes empirical knowledge from previous studies and guides future research by suggesting testable hypotheses. |
| **Important Features** | Intervening variables provide an explanatory link between variables.<br>Good theories are:<br>• *Logical.* They make sense and predictions can be logically deduced.<br>• *Precise.* Predictions about behavior are specific rather than general.<br>• *Parsimonious.* The simplest explanation for a phenomenon is best. |

Based on Table 2.3, Zechmeister, Zechmeister, & Shaughnessy, 2001, p. 29.

*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 must state exactly what a flashbulb memory is and how a flashbulb memory differs from typical memories. The theory would include descriptions of relationships, such as the relationship between degree of emotional involvement and amount remembered (e.g., Talarico & Moore, 2012). 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 is very confident about the (inaccurate) memory (see Neisser & Harsch, 1992). Such was the case in a study that examined people's memory of the September 11 attacks one and two years after the event (Conway, Skitka, Hemmerich, & Kershaw, 2009). Despite a decrease in the accuracy and consistency of their memories over time, participants maintained a high degree of confidence in their flashbulb memory.

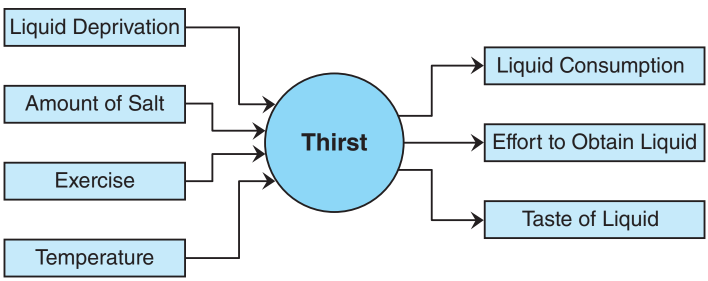
The major functions of a theory are to *organize* empirical knowledge and to *guide* research (Marx, 1963). Even in relatively specific areas of research such as flashbulb memories, many studies have been done. As the scope of a research area increases, so does the number of relevant studies. Scientific 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. 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, click the mouse, or tap a touchpad, you see (and hear) various outcomes on the screen printer, and from the speakers. Yet it isn't your keystrokes and taps that *directly* cause these outcomes; the intervening variable is the "invisible" software applications that serve as a connection between your actions and the outcome on your screen.

Intervening variables are like computer applications. Corresponding to the connection between keystrokes and what you see on your screen, 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" (see Figure 2.6). 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

**FIGURE 2.6** Potential independent variables (left) may influence the intervening variable of "thirst," as measured by potential dependent variables (right).



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 in Figure 2.6 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 increase the acceptability of a theory; unsuccessful tests decrease the theory's acceptability. 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 testable hypotheses. A theory is not likely to fail on the basis of a single test. 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 the use of empirical procedures rather than 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 the 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** 31
**independent variable** 32
**dependent variable** 33
**construct** 33
**operational definition** 34
**validity** 38
**reliability** 38
**correlation** 45
**causal inference** 47
**confounding** 48
**applied research** 49
**basic research** 49
**theory** 50

## Review Questions

**1** For each of the following characteristics, distinguish between the scientific approach and everyday approaches to knowledge: general approach and attitude, observation, concepts, reporting, 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 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 psychologist conducted an experiment to test the hypothesis that individuals embedded in their in-group culture would be less likely to help a stranger. College students were recruited to respond to "a brief survey about their campus experience" near the entrance to the student activity center. The first testing session took place early in the semester. To activate identification with their university (embeddedness), these participants were given a clipboard and asked to write down three things they like about their university. Twenty students were tested. The second testing session took place on two afternoons during the last week of classes at the same location. In this control condition (low-embedded situation), 20 new students were asked to write down three things they plan to do during break.

In each condition, immediately after each participant returned the clipboard to the psychologist, a student research assistant, wearing a sweatshirt with the name of a rival school, walked by the pair and "accidentally" dropped a file containing papers near the participant. The psychologist recorded whether the participant helped pick up the papers. Results indicated that, as predicted, participants in the embedded condition were less likely to help than participants in the control condition. The psychologist concluded that identification with an in-group (embeddedness) causes people to offer less help to a stranger.

**A** Identify the independent variable of interest to the psychologist (and its levels) and the dependent variable.

**B** What potentially relevant independent variable is confounded with the psychologist's independent variable? Explain clearly how the confounding occurred and describe the conclusions that can be made about the effects of embeddedness on helping.

**C** Suggest ways in which the experiment could be done so the psychologist could make a clear conclusion about the effect of identification with an in-group (embeddedness) and helping a stranger.

3 In a widely distributed news report in March 2013, researchers linked 180,000 obesity-related deaths worldwide (including about 25,000 in America) to the consumption of sugary beverages such as soda, energy, and sports drinks. Using 2010 data from the Global Burden of Diseases Study collected by the World Health Organization, the researchers investigated obesity-related deaths due to diabetes, cardiovascular disease, and cancer. They also obtained data for the per-capita consumption of sugary beverages for the countries in the health study. As sugary-beverage consumption increased, the risk of obesity-related deaths increased. The researchers claimed that overall, 1 in 100 deaths of obese people globally is caused by drinking too many sweetened beverages. Prominent nutritionists have claimed that sugary beverages are a major contributor to the obesity epidemic in the United States. These data have been used by some government officials to call for limits on the size of soft drinks (e.g., New York's Bloomberg law).

**A** The researchers claim that consumption of sugary beverages leads to an increased risk of obesity-related death, and argue that limiting sugary-beverage consumption is an important step in reducing obesity-related deaths. What evidence from this summary can be used to meet the conditions necessary for drawing this causal inference and what evidence is lacking?

**B** What sources beyond this summary would you want to check before reaching a conclusion about these findings? (You might begin with www.cnn.com/2013/03/19/health/sugary-drinks-deaths.)

**4** A study was done to determine whether the use of "clickers" as an instructional method would improve students' test performance in an educational psychology class (Mayer et al., 2009). In the clicker class (academic year 2005–2006), students used clickers to respond to multiple-choice questions during lectures. In the paper-and-pencil class (2006–2007), students responded to multiple-choice questions during lectures using a paper-and-pencil format. In the control condition (2004–2005), the instructor did not present multiple-choice questions in lectures. Results for the midterm and final exams indicated that students in the clicker class performed better than students in the paper-and-pencil and control classes. The researchers concluded that the use of clickers during lectures helps students to perform better on tests, and suggested that the clickers help students to engage in appropriate cognitive processing during learning.

**A** What evidence is present in this summary to meet the conditions for a causal inference between the instructional method and students' test performance? What evidence is lacking? (Be sure to describe the three conditions for a causal inference.)

**B** Identify the four goals of the scientific method and explain whether each is met on the basis of findings from 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, positive emotion, 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 caused 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