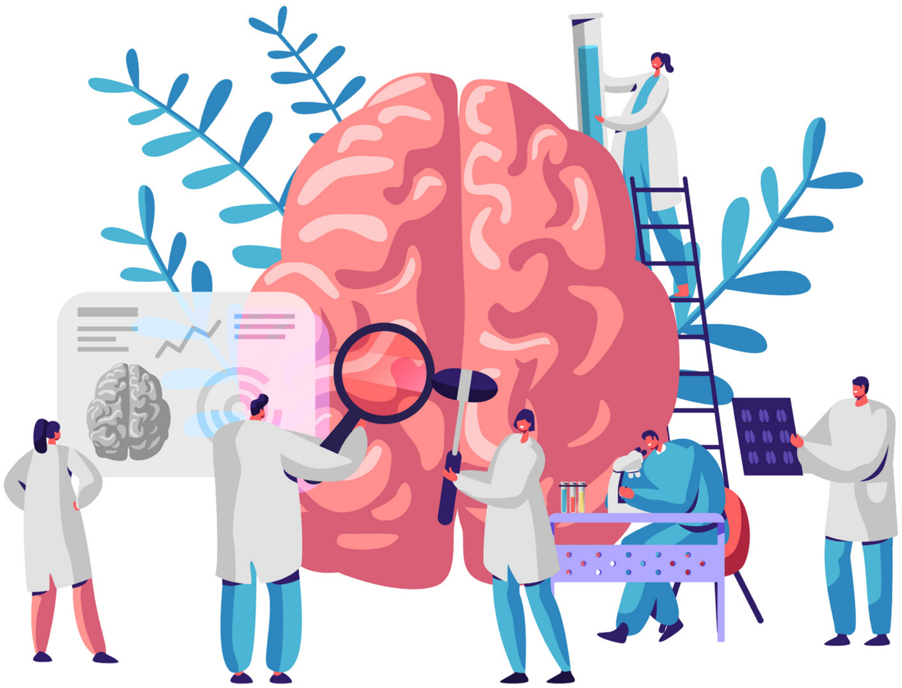
Psychology as a Science

CHAPTER **2**



Source: ivector/Shutterstock.com.

Chapter Outline

Sources of Knowledge
Goals of Scientific Research
Methods of Psychological Research
Statistical Analysis of Research Data
Ethics of Psychological Research

As discussed in Chapter 1, psychology is the science that studies behavior and cognitive processes. It therefore deals with topics of great interest to people, making them also of particular interest to the media. But the media at times are more interested in attracting readers, viewers, and listeners than in the objective reporting of scientific findings. This focus often leads the media to exaggerate or sensationalize research findings. Consider the media's coverage of the supposed effects of the hormone melatonin in the section "Critical Thinking about Psychology."

In discussing psychology as a science, this chapter will answer questions such as these: Why do psychologists use the scientific method? What are the goals of psychological research? How do psychologists employ the scientific method in their research? What techniques do psychologists rely on to analyze their data? And what ethical principles guide psychological research? The answers to these questions will help you appreciate the scientific basis of the issues, theories, research findings, and practical applications presented throughout this book.

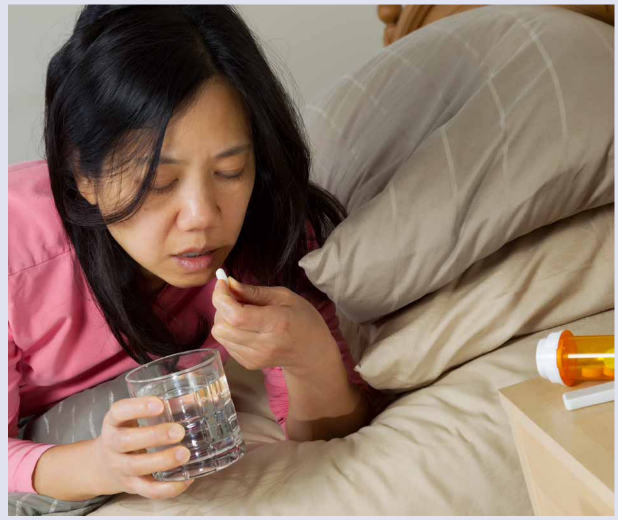
Critical Thinking About Psychology

Does Melatonin Have Beneficial Physical and Psychological Effects?

In November 1995, *Newsweek* magazine's cover story reported a craze inspired by the supposed beneficial physical and psychological effects of a "natural wonder drug," the hormone melatonin (Cowley, 1995). Melatonin, secreted by the pineal gland (located in the center of the brain), was touted in the article as a cure for aging, insomnia, and jet lag. And *Newsweek* was not alone. Reports by magazines, newspapers, radio stations, and television networks across the United States stimulated public excitement about melatonin.

The effects of the media reports were so powerful that many health-food stores could not keep up with consumer demand for melatonin. At the time the *Newsweek* article was published, a book praising the effects of melatonin was third on the *New York Times* best-seller list. Though the craze has subsided since 1995, the media still include periodic reports on the effects of melatonin—and the Internet is brimming with websites that praise the alleged benefits of melatonin, while they just so happen to offer it for sale online.

Should readers have accepted the claims about melatonin's alleged beneficial effects simply because they appeared in a popular magazine that relied mainly on testimonials from people who used or marketed it? Psychologists, being scientists, do not accept such claims unless they are supported by sound scientific research findings that can be replicated. Since the craze, research studies have *indeed* found that melatonin may play a role in countering jet lag (Paul et al., 2010), preventing heart damage in rats and human males after exercise (Ochoa et al., 2011; Veneroso et al., 2009), slowing the aging of mitochondria in mice brains (Carretero et al., 2009), and possibly in treating some early stage cancers (Kubatka et al., 2018). As you read this chapter, you will learn how a psychologist might use the scientific method to conduct an experiment to test the effects of melatonin. But you must first understand the nature of psychology as a science.



Science Requires Critical Thinking

Should we accept media reports as strong evidence for popular claims, such as melatonin's alleged ability to promote sleep, overcome jet lag, and slow the aging process? Scientists require more rigorous standards of evidence than that.

Source: tab62/Shutterstock.com.

Sources of Knowledge

Psychologists and other scientists favor the scientific method as their means of obtaining knowledge, such as understanding the effects of melatonin. To appreciate why science is so vital to our knowledge acquisition, you need to understand the shortcomings of the everyday alternative to the scientific method: *common sense*.

Common Sense

When you rely on common sense, you assume that the beliefs you have obtained from everyday life are trustworthy. Commonsense knowledge has a variety of sources, including statements by recognized authorities, your own reasoning about things, and observations from your personal experience. And, of course, sometimes these things turn out to be true. Many college students view psychology as little more than common sense—until they are presented with examples of how scientific research has demonstrated that some of their commonsense beliefs are false (Osberg, 1993). For instance, your parents, caregivers, or elders may have told you that the keys to a happy life are getting married, having kids, and making money. It turns out that they were right—about two out of three of these things (see the end of this chapter for which two!).

Here is a compelling example of two important themes of psychological science and of this textbook: the frailty of common sense and the intricacy and complexity of gathering

Common Sense

Should you change your answers on multiple-choice tests? You may have heard that you should always stick with your initial answer. Most college students agree with that strategy (Kruger et al., 2005; Merry et al., 2021) and many instructors believe it as well (Merry et al., 2021). You might be surprised then that scientific research has consistently found that students are more likely to change a wrong answer to a correct answer than a correct answer to a wrong one (Benjamin et al., 1984; Merry et al., 2021).

Source: fizkes/Shutterstock.com.

scientific evidence instead. Consider most sports fans' belief in the "hot hand," especially in professional basketball. According to this belief, a player's performance will temporarily improve following a string of successful shots. Announcers will describe a specific player as "hot" or "on fire" or "in the zone," and imply that they should be taking the team's next shots. And yet, despite what we may think we see, a meta-analysis of actual shooting data demonstrates that the vast majority of players are no more likely to make a shot after a series of hits than a series of misses (Avugos et al., 2013). The original "hot hand" research was done in the 1980s based on the shooting records of the Philadelphia 76ers, the Boston Celtics, and Cornell's men's and women's varsity basketball teams (Gilovich et al., 1985). Other analyses of free throws revealed that—contrary to common sense—players did *not* perform better following a "streak" of successes. In fact, players performed at about their base rate following a sequence of successful shots (Koehler & Conley, 2003). Thus, the best predictor of basketball players' next shots is their average performance, not the perception that they have "hot hands." However, an alternative statistical procedure showed that the hot hand myth may not be entirely mythical after all (J. B. Miller & Sanjurjo, 2019). Moreover, there is mounting evidence that a small minority of NBA players can and do get "hot" and occasionally exceed their base rate performance due to mismatches or other nonrandom factors (Pelechrinis & Winston, 2022), so even this research is far more complicated and nuanced than it might at first appear to be.

Nevertheless, we should not automatically discount the possibility that commonsense beliefs might be true. According to Harold Kelley (1921–2003), a leading researcher on human thinking, "discarding our commonsense psychology baggage would require us needlessly to separate ourselves from the vast sources of knowledge gained in the course of human history" (H. H. Kelley, 1992, p. 22). In other words, common sense may inspire scientific research, even though it cannot substitute for it—as in the research study inspired by a major social problem discussed in the "Psychology Versus Common Sense" box.

Science

Because of the weaknesses of common sense and the need for a more objective, self-correcting source of knowledge, scientists prefer the scientific method, which is based on certain assumptions and follows a formal series of steps. Remember, the word *method* is a combination of the Greek words *meta* and *hodos* (see Chapter 1). The fact that the scientific method is the dominant research method in psychology owes much to its origins in 19th-century natural science, particularly physiology, but goes back much further than that as a genuinely global phenomenon.

Origins of Science

Humankind has always been curious, striving to understand why the world works the way it does and linking observation with prediction to discern patterns in the chaotic universe. For example, since prehistoric times, we have observed the skies and learned to predict the seasonal changes in the position of the sun, moon, and stars. Early humans also observed that certain substances (e.g., those derived from plants) could be used to treat disease, and herbal medicines were developed, some of which are still used by modern pharmacotherapy. Science is a special way of generating knowledge about the world that was simultaneously developed in India, China, the Middle East, and South America, so its roots are both ancient and global. The Babylonians, for instance, were among the first peoples to record their systematic observations of the planets and stars (Bynum, 2013). Today's mathematics—the foundation of science—stems from India via the Middle East, whereas China and Egypt employed science to build amazing structures that have stood the test of time, such as the Great Wall and the Pyramids (Bynum, 2013). In Latin America, moreover, there was a "secret" and "endogenous science" being developed centuries before European settlers began to arrive (Saldaña, 2006). Scientific thought flourished during the Ancient Greek civilization ending a few hundred years B.C.E., with Euclid and Pythagoras advancing geometry and Aristotle proposing a view of nature so powerful that it lasted long after his death (Bynum, 2013). However, it was not until the 13th century that much of this scientific work was brought together and formalized in European universities to more closely resemble the science you study at your college or university today.

Assumptions of Science

Scientists share some basic assumptions that guide their thinking about physical reality. Two of the most important of these assumptions are determinism and skepticism.

Determinism and Lawfulness Albert Einstein was fond of saying, "God does not play dice with the universe." In using the scientific method, psychologists and other scientists share his belief that there is order in the universe, meaning that the relationships among events are lawful rather than haphazard. In looking for these predictable relationships, scientists also share the assumption of **determinism**, which holds that every event has physical, probably measurable, causes. Determinism therefore rules out free will and supernatural influences as causes of behavior.

determinism The assumption that every event has physical, potentially measurable, causes.

Yet, as pointed out more than a century ago by William James, scientists might be committed to determinism in conducting their research, while being tempted to assume the existence of free will in their everyday lives (Immergluck, 1964). Despite centuries of philosophical debate, neither side of the determinism versus free will debate has won. This controversy is one that neither psychologists nor philosophers have been able to resolve, though some still try by resorting to soft determinism, which asserts that determinism generally governs events though at times we can impose free will on our actions (Clarke, 2010). What do you think? Did you choose to read that last sentence?

Skepticism and Critical Thinking Aside from assuming that the universe is an orderly place in which events—including behaviors—are governed by determinism, scientists today, like René Descartes and Francis Bacon before them (see Chapter 1), insist that open-minded **skepticism** is the best intellectual predisposition when judging the merits of any claim. Such skepticism requires the maintenance of a delicate balance between cynicism and gullibility. As Mario Bunge, a leading philosopher of science, has noted, skeptics "do not believe anything in the absence of evidence, but they are willing to explore bold new ideas if they find reasons to suspect that they have a chance" (Bunge, 1992, p. 380). This attitude requires supportive evidence before accepting any claim. The failure to maintain a skeptical attitude leads to the acceptance of phenomena that lack sufficient scientific support (R. E. Bartholomew & Radford, 2003), which could range from ESP and psychic mediums to fad diets and inert pills, such as homeopathy, that do not work.

skepticism An attitude that doubts all claims not supported by solid research evidence.

Skepticism also is important in psychology because many psychological truths are tentative, in part because psychological research findings may depend on the times and places in which the research takes place. What generally is true of human behavior in one era or culture might be false in another. For example, gender differences in behavior in Western cultures have changed dramatically over the past few decades, and gender

Psychology Versus Common Sense

Can We Reliably Detect When Someone Is Intoxicated?

In the landmark 1961 *Zane* decision, a New Jersey court stated, "Whether the man is sober or intoxicated is a matter of common observation not requiring special knowledge or skill" (Langenbucher & Nathan, 1983, p. 1071). This assumption is an important one because state laws in the United States, based on the commonsense belief that intoxication is easily detected, hold people such as party hosts and bar owners legally responsible for the actions of people who become intoxicated at their homes or businesses. The ability for others to detect intoxication was tested in a scientific study by researchers James Langenbucher and Peter Nathan (1983). Although this research was conducted decades ago, the findings and implications are still very relevant.

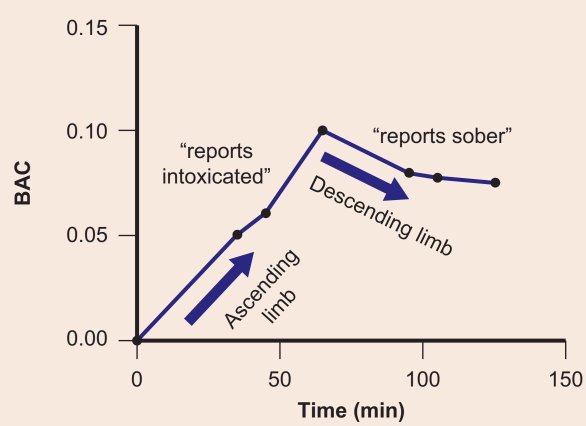
The researchers asked 12 bartenders, 49 social drinkers, and 30 police officers to observe drinkers and judge whether they were legally drunk or sober (without a breath test or blood test). The drinkers in each case were two male and two female young adults. Each drinker consumed one of three drinks: tonic water (**the control**), moderate doses of vodka (but not enough to become legally drunk), or high doses of vodka (enough to become legally drunk). A breathalyzer ensured that the desired blood-alcohol concentrations (BAC) were achieved for participants in the two vodka conditions.

The bartenders observed their participants being interviewed in a cocktail lounge. The social drinkers observed their interviews in the researchers' laboratory at Rutgers University. And the police officers observed their participants in a simulated nighttime roadside arrest in which they were given 3 minutes to determine whether the motorist they had pulled over was intoxicated or sober.

Langenbucher and Nathan found that the observers correctly judged the drinkers' level of intoxication only 25% of the time. Not a single legally intoxicated person was identified as such by a significant number of the observers. Of the 91 people who served as judges, only five were consistently accurate—and all of them were members of a State Police special tactical unit for the apprehension of drunk drivers. Those five police officers had received more than 90 hours of training in the detection of drunkenness. The results implied that, without special training, even people with extensive experience in observing drinkers might be unable to determine whether a person is legally drunk or sober.

The social implication of these findings is that common sense is wrong in the assumption that people with experience in observing drinkers can detect whether someone is intoxicated. We are even more confident in the findings of this study because they were supported by the results of a different experiment conducted by a different researcher, using different participants, in a different setting (Brick & Carpenter, 2001)—the fancy word for this is *replication*. Perhaps bartenders, police officers, and habitual party givers should obtain special training similar to that given to the five police officers who performed well in the initial study. Or perhaps those five police officers had exceptional olfactory abilities! Either way, would you sign up for that training?

Simulated Driving Performance Impaired on Both Ascending and Descending Limbs of BAC



The Detection of Intoxication

Scientific research contradicts the commonsense belief that we can easily detect when someone is legally drunk. The ability to detect your *own* intoxication has also been scientifically studied by researchers. Social-drinking adults were given doses of alcohol to induce intoxication. Alcohol first increases in your bloodstream (ascending limb) and then over time decreases (descending limb) through metabolism. Researchers simulated driving performance and found that subjects made risky decisions to drive and the same driving impairments in the descending limb as in the ascending limb despite reporting feeling "sober" (Weafer & Fillmore, 2012). So scientific research also contradicts the commonsense belief of feeling sober enough to drive. Please find a safe ride and take the keys from your friends when necessary!

Source: (Photo) PBXStudio/Shutterstock.com; (line art) Adapted from Weafer & Fillmore (2012).

differences observed in Western cultures might be unlike those in non-Western cultures. More than two decades before the sociocultural perspective (see Chapter 1) achieved widespread acceptance in North American psychology, Anne Anastasi (1908–2001), in her presidential address to the American Psychological Association, showed foresight in urging psychologists not to confuse their ethnocentric personal beliefs and values with scientific knowledge (Anastasi, 1972).

Skepticism is valuable not only for scientists but for all of us in our everyday lives, as we evaluate information presented in academic courses, media reports, and websites. Skepticism also is the basis of *critical thinking*—the systematic evaluation of claims and assumptions. Students who major in psychology tend to become particularly adept at critical thinking by their senior year (T. J. Lawson, 1999). The following steps in critical thinking will serve you well as you evaluate claims encountered in your everyday life. Critical thinking involves asking yourself these seven questions about any claim that you see, hear, or read (Burke et al., 2014):

1. What are you being asked to believe or accept?
2. What evidence is available to support the claim?
3. What alternative ways are there to interpret the evidence?
4. How would you rate all the evidence/alternatives on a 0–10 scale based on validity/strength? This is a key step. Scientific evidence, especially if it has been replicated, should garner higher ratings than what your friends, parents, caregivers, or elders told you (with all due respect to them).
5. What assumptions or biases came up when answering questions 1–4 (e.g., using intuition/emotion, authority, or personal experience rather than science)?
6. What additional evidence would help you evaluate the alternatives?
7. What conclusions are most reasonable or likely here? This is not what you want to believe, but what the preponderance of the evidence suggests is true.

Effectively using any or all of these seven questions can help you live better and be less likely to fall prey to misinformation or advertisers who want to sell you on their ideas or their products for their own (but not your) benefit.

Steps in Conducting Scientific Research

Critical thinking is a strong way of deciding what is true and lies at the heart of the scientific method; science has been described as formalized critical thinking, although it goes even beyond critical thinking in its pursuit of acquiring new knowledge. Because psychologists are skeptical about claims not supported by research findings, they employ the **scientific method** as the best means of gaining knowledge and the only one throughout human history that excels at making *predictions*. For instance, what do you do when you have a trip planned and you want to know the weather? One of your textbook authors used to ask his grandmother, who invariably (though often erroneously) stated that "it is going to be beautiful tomorrow." He eventually found out that she just wanted it to be nice and he now relies exclusively on online weather reports from the government, which are becoming increasingly accurate (Fry, 2019). Though scientists vary in their approach to the scientific method, ideally, they follow a formal series of steps:

scientific method A source of knowledge based on the assumption that knowledge comes from the objective, systematic observation and measurement of particular variables and the events they affect.

Step 1: Provide a *rationale* for the study. The scientist identifies the problem, reviews the relevant research literature, decides on the research method to use, and states the research **hypothesis**. A hypothesis (from the Greek word for "supposition") is a testable prediction about the relationship between two or more events or characteristics.

hypothesis A testable prediction about the relationship between two or more events or characteristics.

Step 2: Design the study. Scientists, using critical thinking and specific tools for experimental design, must carefully plan out how the study will test their hypothesis and rule out any possible alternatives or confounds. We will have much more to say about this later in the chapter.

Step 3: Conduct the study. The scientist carries out the research procedure and collects data.

Step 4: Analyze the data. The scientist usually uses mathematical techniques called *statistics* and discusses the implications of the research findings.

Step 5: Communicate the research findings. The scientist may present the research study at a professional meeting and/or publish an article describing the study in a professional journal. In doing so, the scientist includes the rationale for the research, the exact method that was used, the results of the research, and a discussion of the implications of the results.

Step 6: Replicate the study. **Replication** involves repeating the study, exactly or with some variation. Either the original researcher or, even better, other researchers may replicate the study. Successful replications of research studies strengthen confidence in their findings.

replication The repetition of a research study, usually with some alterations in its methods or setting, to determine whether the principles derived from that study hold up under similar circumstances.

These steps were used by psychologist Donn Byrne and his colleagues (1970) in a classic research study of interpersonal attraction: Do opposites attract or do birds of a feather flock together? In this study, the problem concerned the relationship between interpersonal similarity and interpersonal attraction. After reviewing the research literature relevant to the problem, Byrne decided to conduct a *field experiment* that studied college students in a real-life setting instead of in a laboratory. In fact, his experiment was a replication study to determine whether the results of previous laboratory studies on the effects of attitude similarity on interpersonal attraction would generalize to the real world.

Based on his review of the research literature, Byrne hypothesized that heterosexual men and women with similar attitudes would be more likely to be attracted to each other than would those with dissimilar attitudes. Byrne had his research participants complete a 50-item questionnaire that assessed their attitudes as part of a computer-dating service. He told them that their responses would be used to pair them with a student who shared their attitudes. But the students were actually paired so that some partners were similar in attitudes and others were dissimilar. Their similarity on the questionnaire provided a concrete definition of "similarity." The 44 heterosexual couples, selected from 420 volunteers, then were sent to the student union for a 30-minute get-acquainted date. Several weeks later, participants were asked to rate their partners, which provided Byrne with his research data.

Like almost all researchers, Byrne used statistics to summarize his data and to determine whether his hypothesis was supported. In this case, Byrne found that the data did support the hypothesis. Partners who were similar in attitudes were significantly more likely to recall each other's name, to have talked with each other since the date, and to desire to date each other again. Thus, in this study, the use of the scientific method found that birds of a feather tend to flock together.

Byrne communicated his findings by publishing them in a professional journal and shared his findings by presenting them at a research conference. Even undergraduate psychology researchers can present the results of their research studies at psychology research conferences held each year. If you are in the United States, be sure to check out all the regional affiliated American Psychological Associations conferences such as the New England Psychological Association (NEPA), the Rocky Mountain Psychological Association (RMPA), and many more.

It is valuable to communicate your research findings because one of the main differences between science and pseudoscience is that science is always changing, being updated with new information and knowledge, which is why this textbook is already in its 9th edition. By contrast, pseudoscience remains fixed; for example, the assessment of your personality by the bumps in your head (phrenology, discussed in Chapter 3), treating respiratory conditions with homeopathic preparations (D. King et al., 2021), or the analysis of your handwriting (graphology) never change or update themselves and none has garnered any empirical (scientific) support.

In the past decade, seminal papers in biomedical science and psychology claimed that only about half of the findings in these two fields were replicated in future research (L. P. Freedman et al., 2015; Open Science Collaboration, 2015). You can look at this glass as half empty or half full, but there are clear lessons to be learned here: Namely, that science is dynamic and ever-changing, so replication is vital to get us as close to the truth as we can. You will see throughout this textbook that some “classic” psychology findings turned out to be correct, whereas others did not despite their early promise. For example, even Byrne’s research on the similarity-attraction effect varies across cultures, being stronger in the United States than in Japan (S. J. Heine et al., 2009). Because of this, a new focus in psychology has been the notion of “open science,” including preregistering research studies and making all materials and procedures available so that other researchers can attempt to replicate the findings (Munafò et al., 2017). Truth—and science—takes time.

Section Review: Sources of Knowledge

1. What are the basic assumptions of science?
2. What is critical thinking?
3. What are the formal steps in the scientific method?

Goals of Scientific Research

In conducting their research, psychologists and other scientists share common goals. They pursue the goals of description, prediction, control, and explanation of behavior and cognitive processes (C. D. Green & Powell, 1990).

Description

To a scientist, description involves noting the observable characteristics of an event, object, or individual. For example, we might note that participants who take daily doses of melatonin sleep longer. Psychologists, following in the intellectual tradition of Francis Bacon discussed in Chapter 1, are systematic in what they describe. Instead of arbitrarily describing everything that they observe, they describe only things that are relevant to their research topic. Thus, good observational skills are essential to psychologists. The need to be systematic in what you describe is expressed well in a statement about criminal investigations made by the fictional detective Sherlock Holmes to his friend Dr. Watson:

> *A fool takes in all the lumber [facts] that he comes across, so that the knowledge which might be useful to him gets crowded out, or at best is jumbled up with a lot of other things . . . It is of the highest importance, therefore, not to have useless facts elbowing out the useful ones.* (A. C. Doyle, 1930)

measurement The use of numbers to represent events or characteristics.

In science, descriptions must be more than systematic; they must be precise. Precise descriptions are concrete rather than abstract, and typically involve **measurement**, which is the use of numbers to represent events or characteristics. According to Francis Galton, one of the pioneers of psychology, “Until the phenomena of any branch of knowledge have been submitted to measurement . . . it cannot assume the status and dignity of a science” (quoted in Cowles, 1989, p. 2). Thus, describing a friend as “generous” would be acceptable in everyday conversation but would be too imprecise for scientific communication. Think about how hard it would be to describe your friends or family scientifically!

operational definition The description of behaviors or qualities in terms of the procedures used to measure or produce them. Creating an operational definition is *operationalization*.

Scientists solve this problem by using **operational definitions**, which define behaviors or qualities in terms of the procedures used to measure or produce them (Feest, 2005). Donn Byrne did this when he defined *similarity* according to participants’ responses to a questionnaire in his study of interpersonal attractiveness. More than a century ago, Galton, in studying audience behavior at plays and lectures, operationally defined *boredom* by recording the number of fidgets by audience members. You might operationally define *generous* as “donating more than 5% of one’s earnings to charity.” And a common operational definition of being *legally drunk* in the United States is “a blood-alcohol concentration of at least 0.08%." Though operational definitions are desirable, psychologists sometimes find it difficult to agree on acceptable ones. For example, a series of journal articles argued about how best to operationally define "suicide attempt" (Kidd, 2003). Another example is the operational definition of *anxiety*. The definition could include observable measures such as sweating palms, but it could also include withdrawal from a situation or setting. One benefit of an operational definition is that it promotes more precise communication among scientists.

Prediction

Psychologists are not content just to describe behavior and events but also make predictions in the form of hypotheses about changes in behavior, cognitive experiences, or physiological activity. A hypothesis is usually based on a **theory**, which is a set of statements that summarize and explain research findings and from which research hypotheses can be derived. For example, Sigmund Freud's theory of psychoanalysis integrates many of his observations of the characteristics of people suffering from psychological disorders. Theories provide coherence to scientific research findings and suggest applications of research findings, making science more than the accumulation of isolated facts (Kukla, 1989). But, in order to be useful, theories need to lead to accurate predictions, something that many modern psychologists would say Freud's theory lacks.

theory An integrated set of statements that summarizes and explains research findings and from which research hypotheses can be derived.

As you might guess, human (or animal) behavior can be especially challenging to predict, prompting the notion that "soft [social] sciences are often harder than hard [physical] sciences" (J. Diamond, 1987, p. 35). Moreover, such scientific predictions are more accurate when applied to multiple participants rather than when applied to a single case. For example, your automobile insurance company can more accurately predict the percentage of people in your age group who will have an accident this year than it can predict whether you will have one. Likewise, although melatonin might prove effective in treating insomnia for most people, we would be unable to predict with certainty whether you would benefit from it. Psychologists sometimes muse that we can predict the behavior of thousands of people but not any one of them, which is why it is difficult to predict the next school or mass shooting (see Chapter 1).

Psychology has nothing to apologize for in being limited to probabilistic prediction because this situation is no different in the other sciences, which likewise are limited to making predictions that are not always correct (Hedges, 1987). Your physician might prescribe an antibiotic that, based on medical research, is effective 96% of the time in treating pneumonia, but they cannot guarantee that it will cure *your* pneumonia. Vaccines such as those for COVID-19 work for most people (Baden et al., 2021), but perhaps you know someone for whom they did not work so well. Similarly, flood forecasters know that regions along certain rivers are more likely to flood (Reggiani & Weerts, 2008) and earthquake forecasters know that regions along geological faults are more likely to experience earthquakes (Molchan & Keilis-Borok, 2008), but they cannot accurately predict well in advance the day, or even the year, that a flood or an earthquake will occur in a given region. In the same vein, in regard to interpersonal attraction, people with similar attitudes will probably—but not always—be more attracted to each other than people with dissimilar attitudes. Alas, we cannot predict with certainty whether your next date with a person who is similar to you will go well.

Control

Psychologists go beyond describing and predicting changes in behavior or cognitive processes to influence those changes by controlling factors that affect them. The notion of *control* is used in two ways (Cowles, 1989). First, as you will read later in the chapter, control is an essential and critical ingredient in the conduct of experiments. Second, psychologists may apply research findings to the control of behavior in everyday life (L. D. Smith, 2002). Thus, melatonin might be prescribed to control insomnia or circadian shifts by promoting sleep, anxiety may be treated by cognitive-behavioral therapy (CBT), and you might be advised to find romance by seeking someone who shares your values and interests.

Psychologists seek to help individuals gain control over phenomena as diverse as psychological disorders (Mansell & Carey, 2009) and type 2 diabetes (A. B. Gonzalez et al., 2011).

Explanation

The ultimate goal of psychology, however, is explanation—the discovery of the causes of behaviors and cognitive processes. If it is demonstrated that people who ingest melatonin consistently overcome insomnia, the next step might be to explain how and where melatonin affects the brain to trigger sleep (remember, it is produced in the pineal gland; see "Critical Thinking About Psychology" earlier in this chapter). Likewise, even though we know that attitude similarity promotes interpersonal attraction, we still would need to explain why we prefer people who have similar attitudes.

As discussed in Chapter 1, psychologists' favored perspectives determine where they look for explanations of psychological phenomena, such as psychological disorders (Lam et al., 2005). Psychologists who favor the cognitive, humanistic, or psychoanalytic perspective will look for causes in the mind. Psychologists who favor the behavioral perspective will look for causes in the environment. Psychologists who favor the biopsychological perspective will look for causes in the brain or hormonal system. And psychologists who favor the sociocultural perspective will look for causes in the social or cultural context of the event. And, of course, none of these are correct on their own—every event (such as the school shooting described in Chapter 1) has multiple causes and therefore multiple explanations.

Section Review: Goals of Scientific Research

1. Why do scientists use operational definitions?
2. In what way are psychology and other sciences probabilistic?
3. What is the nature of scientific explanation in psychology?

Methods of Psychological Research

Given that psychologists favor the scientific method as their primary source of knowledge, how is this method employed in conducting research? And once data are collected, how can we make sense of it? As shown in Table 2-1, psychologists use research methods that permit them to describe, predict, control, or explain relationships among variables.

Descriptive Research

descriptive research Research that involves the recording of behaviors that have been observed systematically.

Descriptive research means that researchers simply record what they have systematically observed. Descriptive research methods include *naturalistic observation*, *case studies*, *surveys*, *psychological testing*, and *archival research*.

TABLE 2-1 The Goals and Methods of Psychology

Goal	Research Method	Relevant Question
Description	Descriptive	What are its characteristics?
Prediction	Correlational	How likely is it?
Control	Experimental	Can I make it happen?
Explanation	Experimental	What causes it?

Naturalistic Observation

In **naturalistic observation**, people or animals are observed behaving in their natural environment. Researchers who use naturalistic observation study topics as diverse as the ability to find where one has parked one's car (Lutz et al., 1994), peer reactions to bullying on school playgrounds (D. L. Hawkins et al., 2001), and factors related to smiling during group interactions (Mehu & Dunbar, 2008). To make sure that their observations represent natural behavior, observers refrain as much as possible from influencing the individuals they are observing. In other words, the observer remains unobtrusive. If you were studying the eating behavior of students in your school cafeteria, you would not announce your intention over a loudspeaker. Otherwise, your participants might behave unnaturally; a person who normally gorged on cake, ice cream, and chocolate pudding for dessert might eat fruit instead.

naturalistic observation The recording of the behavior of people or animals in their natural environments, with little or no intervention by the researcher.

Naturalistic observation also is useful in studying animal behavior. Some of the best-known studies employing naturalistic observation were conducted by Jane Goodall, who spent decades observing chimpanzees in Gombe National Park in Tanzania (Crain, 2009). To prevent newly encountered chimpanzees from acting unnaturally because of her presence, Goodall spent her initial observation periods letting them get used to her.

The study of animal behavior in the natural environment, as in Goodall's research, is called **ethology**. One of the advantages of an ethological approach is the potential discovery of behaviors not found in more artificial settings, such as zoos and laboratories. Goodall reported observations concerning mundane chimpanzee behavior, such as "fishing" for ants with sticks (O'Malley et al., 2012) and observations that have not been made in captivity, including cannibalism, infanticide, and the unprovoked killing of other chimpanzees (Goodall, 1990). But researchers who use naturalistic observation, like those who use other research methods, must not be hasty in generalizing their findings. Even Jane Goodall's observations must be qualified: The behavior of the Gombe chimpanzees differs from the behavior of chimpanzees in the Mahale Mountains of western Tanzania, where female chimpanzees hunt more often (Takahata et al., 1984). A more recent researcher at the intersection of ethology and behavioral neuroscience is Robert Sapolsky. Sapolsky studied the social behaviors of the same baboons in the wild for 25 years and found that the stress hormone, cortisol, fluctuates the most in male dominant baboons (Sapolsky, 1992). His work on dominance hierarchies in nonprimates is translatable to human physiological mediators of socioeconomic status and psychological health.

ethology The study of animal behavior in the natural environment.

Naturalistic observation cannot determine the causes of the observed behavior because there are simply too many factors at work in a natural setting. So you could not determine


Naturalistic Observation

Jane Goodall's naturalistic observations of chimpanzees in the wild have contributed to our understanding of their everyday habits, many of which had never been observed in zoos or laboratories. Robert Sapolsky's naturalistic observations of baboons through a more modern lens have contributed to understanding stress hormones, physiology, and health.

Source: Tinseltown/Shutterstock.com; Stephen Lew/Shutterstock.com.

why female chimpanzees hunt more in one part of Tanzania than in another. Is it due to differences in prey, in climate, in topography, or in another factor or some combination of factors? It would be impossible to tell just by using naturalistic observation. Modern ethology is interested in both observational studies and experimental studies.

Case Study

case study An in-depth study of an individual.

Another descriptive research method is the **case study**—an in-depth study of a person, typically conducted to gain knowledge about a particular psychological phenomenon that is relatively rare, such as the pathological hoarding of items (Koretz & Gutheil, 2009). The case-study researcher obtains as much relevant information as possible about a host of factors, perhaps including the person's thoughts, feelings, life experiences, and social relationships. The case study often is used in clinical studies of people suffering from psychological disorders. In fact, Sigmund Freud based his theory of psychoanalysis on data he obtained from clinical case studies, which are still a staple of psychoanalytic research (Midgley, 2006).

Most recently, the case study method has been used to gain insight into factors related to a rash of student shootings of their teachers and classmates (see Chapter 1). Researchers conducted case studies of 15 shooting incidents between 1995 and 2001 to examine the possible role of social rejection. Ostracism, bullying, or romantic rejection was present in all but two of the cases. The shooters also tended to have one or more of the following three risk factors: an interest in guns or explosives, a fascination with death or Satanism, and/or psychological problems involving depression, impulse control, or sadistic tendencies (Leary et al., 2003).

Because a person's behavior is affected by many variables, the case study method cannot determine the particular variables that *caused* the behavior being studied. Though it might seem reasonable to assume that the shooters' experiences of rejection caused them to lash out at their teachers and fellow students, that assumption might be wrong. Other factors, unrelated to social rejection, might have caused the violence. It is even conceivable that the shooters' peers rejected them only *after* discovering their fascination with death, Satanism, or guns.

Another shortcoming of the case study is that the results of a single case study, no matter how dramatic, cannot be generalized to all people. Even if the shooters lashed out at their teachers and peers in response to social rejection or bullying, other people who commit violent acts might not have experienced similar rejection—and, of course, many people experience social rejection but do not commit acts of violence. And as you will learn in Chapter 14, numerous studies have shown that both biopsychological and psychosocial factors play a role in violence.

Survey

survey A set of questions related to a particular topic of interest administered to a sample of people through an interview or questionnaire.

When psychologists wish to collect information about behaviors, opinions, attitudes, life experiences, or personal characteristics of many people, they use a descriptive research method called the survey. A **survey** asks participants a series of questions about the topic of interest, such as product preferences or political opinions. Surveys deal with topics as varied as factors related to condom use (S. E. French & Holland, 2013), purchasing habits of students using school vending machines (D. Rose, 2011), occupational stress experienced by university professors (Slišković et al., 2011), and psychological symptoms related to video-game dependency (Rehbein et al., 2010). Surveys commonly are in the form of personal interviews or written questionnaires—sometimes presented online, as in product marketing surveys.

You probably have been asked to respond to several surveys in the past year, whether enclosed in the "You May Have Already Won!" offers that you receive in the mail or conducted by your university or college to get your views on campus policies. If we were surveying you about your recent surveys, we might ask you scaled questions about how many surveys you completed and how much you enjoyed them.

The prevalence of surveys, and the potential annoyance they may induce, is not new. More than a century ago, William James (1890/1981) was so irritated by the seeming

Surveys

The Internet is a fruitful area for conducting survey research. Some advantages for researchers are increased access to unique populations and saving time and money. As technology has evolved (away from mailings and telephone surveys), so have the methods for recruiting participants. Amazon's Mechanical Turk (MTurk) has transformed survey research and allows for researchers to use crowdsourcing as a platform. Some disadvantages are the lack of random sampling and demographic drawbacks because Internet access might not be available to all potential participants.

Source: Andrey_Popov/Shutterstock.com; Koshiro K/Shutterstock.com.

omnipresence of surveys that he called them "one of the pests of life." Today, the most ambitious of these "pests" is the U.S. census, which is conducted every 10 years. Others you might be familiar with include the Gallup public opinion polls, the Quinnipiac Poll, the Pew Research Center surveys, and Harris polls.

High-quality surveys use clearly worded questions that do not bias respondents to answer in a particular way. But surveys are limited by respondents' willingness to answer honestly and by *social desirability*—the tendency to give socially appropriate responses. You can imagine the potential effect of social desirability on responses to surveys on delicate topics such as child abuse, drug use, academic cheating, or sexual practices.

Still another issue to consider in surveys is the effect of sociocultural differences between test takers. You certainly are familiar with questionnaires that ask you to respond on a Likert-type scale from, say, 1 to 7, with 1 meaning "strongly agree" and 7 meaning "strongly disagree." A cross-cultural study of high school students found that they differed in the degree to which they were willing to use the extreme points on scales like this. Students from Japan and Taiwan were more likely to use the midpoint than were students from Canada and the United States. This finding might be attributable to the greater tendency toward individualism in North American cultures and the higher collectivism (the principle of giving a group priority over each individual in it) often present in East Asian cultures (C. Chen et al., 1995). Consequently, survey researchers who use these kinds of scales should consider the cultural backgrounds of their participants when interpreting their findings.

Because of practical and financial constraints, surveys rarely include everyone of interest, even the newer crowdsourcing type of recruitment. Instead, researchers administer a survey to a **sample** of people who represent the target **population**. In conducting a survey at your school, you might interview a sample of 100 students. But for the results of your survey to be generalizable to the entire student population at your school, your sample must be representative of the student body in age, gender, ethnicity, and any other relevant characteristics. Generalizable results are best achieved by **random sampling**, which makes each member of the population equally likely to be included in the sample. Failure to achieve a random sample of respondents might produce bias because those selected to participate might be different from those not selected to participate in regard to the topic of the survey (Menachemi, 2011).

The need for a sample to be representative of its population was dramatically demonstrated in a notorious poll conducted by the *Literary Digest* during the 1936 U.S.

sample A group of participants selected from a population.

population A group of individuals who share certain characteristics.

random sampling The selection of a sample from a population so that each member of the population has an equal chance of being included.

Random (nonbiased) sampling from a population ensures accurate predictions.
Source: petrroudny43/Shutterstock.com.

presidential election. Until then, the *Literary Digest's* presidential poll, based on millions of responses, had accurately predicted each presidential election from 1916 through 1932. In 1936, based on that poll, the editors predicted that Alf Landon, the Republican candidate, would easily defeat Franklin Roosevelt, the Democratic candidate. Yet Roosevelt defeated Landon in a landslide.

What went wrong with the poll? Evidently, the participants included in the survey were a *biased sample*, not representative of those who voted. Many of the participants were selected from telephone directories or automobile registration lists in an era—the Great Depression—when few people had telephones or automobiles. Because Republican candidates attracted wealthier voters than did Democratic candidates, people who had telephones or automobiles were more likely to favor the Republican (Landon) over the Democrat (Roosevelt). The previous polls did not suffer from this bias because economic differences among voters did not significantly affect their party allegiances until the 1936 election. More recently, as polls and their methodologies have expanded, predicting election results has become big business. In the weeks leading up to the November 2016 election, polls across the United States predicted an easy sweep for Democratic nominee Hillary Clinton. Media outlets and pollsters took the heat for failing to project a victory for Donald Trump. A similar 2.5-percentage point polling error in Trump's support was made again in 2020, though Biden still won that election. Some pollsters believe that the issue remains the lack of a representative sample, as people may be more likely to pick up the phone if they score high in the domain of social trust, which has been found to be higher among Democrats (Dickie, 2020).

Psychological Testing

psychological test A formal sample of a person's behavior, whether written or performed.

A widely used descriptive research method is the **psychological test**, which is a formal sample of a person's behavior, whether written or performed. The advantage of good tests is that they help us make less-biased decisions about individuals. There are tests of interests, attitudes, abilities, creativity, intelligence, and personality. Psychological testing has a variety of uses, including helping to decide child custody in divorce cases (Hagan & Hagan, 2008), determining law-enforcement leadership potential (Miller et al., 2009), and assessing the relationship of environmental lead exposure to cognitive, perceptual, and motor performance (Kmiecik-Małecka et al., 2009). As noted by Anne Anastasi (1985), who was an influential authority on psychological testing for several decades (Hogan, 2003), a useful test reflects important principles of test construction: *standardization*, *reliability*, and *validity*.

standardization 1. A procedure ensuring that a test is administered and scored in a consistent manner.
2. A procedure for establishing test norms by giving a test to large samples of people who are representative of those for whom the test is designed.

norm A score, based on the test performances of large numbers of participants, that is used as a standard for assessing the performances of test takers.

There are two major aspects of **standardization**. The first ensures that the test will be administered and scored in a consistent manner. In giving a test, all test administrators must use the same instructions, the same time limits, and the same scoring system. If they do not, test takers' scores might misrepresent their individual characteristics. The second establishes **norms**, which are the standards used to compare the scores of test takers. Without norms, a score on an intelligence test would be a meaningless number. Norms are established by giving the test to samples of hundreds or thousands of people who are representative of the people for whom the test is designed. If a test is to be used in North America, samples might include representative proportions of genders; people from all ethnic groups; lower-, middle-, and upper-class individuals; and urban, rural, and suburban dwellers. Standardized norms have been established for tests that measure things such as factors involved in developmental changes in attention in children (Vakil et al., 2009) and intelligence test scores of American and Canadian children (Reddon et al., 2007).

The use of testing norms became popular in North America in the early 20th century, in part because of the introduction of the Stanford-Binet Intelligence Scale in 1916 by Lewis Terman. In one case, Terman (1918) used the scale's norms to prevent the execution of a young man with intellectual developmental disorder who had committed a heinous murder. The man's score on the Stanford-Binet indicated that his mental age was equivalent to that of a 7-year-old child. Terman testified as a defense witness in opposition to the prosecution's expert witness, who claimed that the young man could perform various activities that only an adult could perform. But the expert witness presented no evidence, only his personal opinion. Terman convinced the jury, using his intelligence scale's norms as objective evidence, and the death penalty was ruled out (Dahlstrom, 1993). The importance of standardized testing for grade school placement, college admissions, and other purposes increased throughout the 20th century (see Chapter 10). Today, standardized testing to ensure that children are progressing satisfactorily in school is mandatory in the United States under the widely publicized No Child Left Behind legislation (Mattai, 2002). That law was replaced in 2015 by the Every Student Succeeds Act, which relies less on standardized tests but still requires states to report on the progress of traditionally underserved kids (A. M. I. Lee, 2015).

In addition to standardized, an adequate psychological test also must be *reliable*. The **reliability** of a test is the degree to which it gives consistent results over time and across administrators. Suppose you took an IQ test and scored 105 (average) one month, 62 (intellectual developmental disorder) the next month, and 138 (gifted) the third month. Because your level of intelligence would not fluctuate that much in 3 months, the test is clearly unreliable. Likewise, you would doubt the reliability of a test that produced different results depending on who administered it.

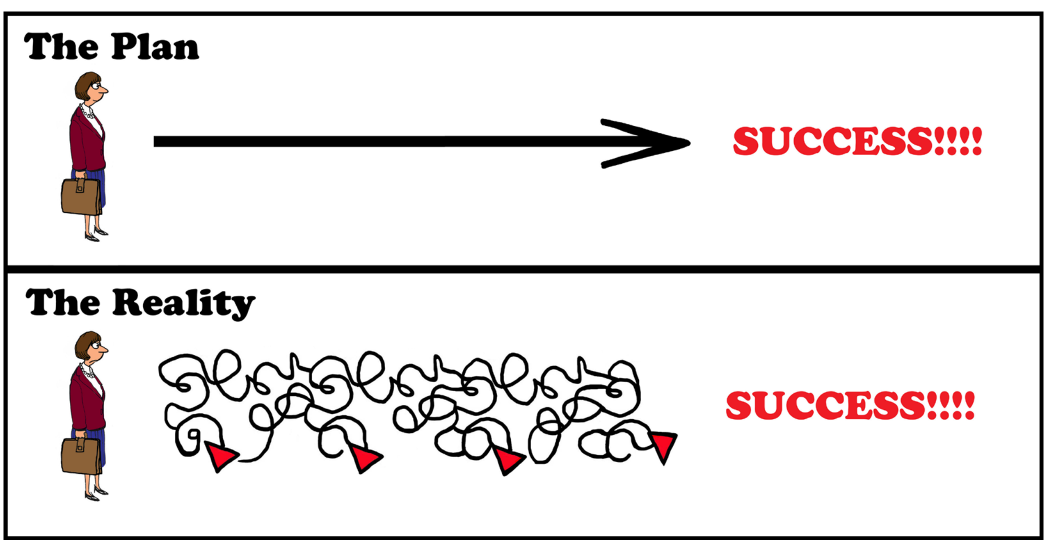
reliability The extent to which a test gives consistent results.

A reliable test would be useless if it were not also valid, however. **Validity** is the extent to which a test measures what it is supposed to measure. An important kind of validity, *predictive validity*, indicates that the test accurately predicts behavior related to what the test is designed to measure. A test of mechanical ability with strong predictive validity would accurately predict who would perform better as an automobile mechanic.

validity The extent to which a test measures what it is supposed to measure.

One of the first studies of the predictive validity of a formal test was conducted by Francis Galton, who collected the civil service exam scores of hundreds of English men who had taken the test in 1861 and their salaries 20 years later. Galton concluded that the exam had good predictive validity, in that those who had scored higher earned higher salaries than did those who had scored lower. More recently, a large-scale review of the predictive validity of the Graduate Record Examination (GRE) found that it is a valid predictor of graduate school performance, as measured by first-year grade-point average (Kuncel et al., 2001). That is, those who score high on the GRE tend to do better in graduate school than those who score low on it.

The sociocultural perspective has inspired greater interest in assessing the extent to which psychological tests, typically developed in North America, have cross-cultural reliability and validity. Tests that have shown cross-cultural reliability and validity include the Portuguese version of the Dental Anxiety Scale (L. W. Hu et al., 2007), the Japanese version of the Social Phobia Inventory (Nagata et al., 2013), the Turkish version of the



How Do You Plan to Enter the Workforce?

Nearly 40% of students majoring in psychology seek full-time employment upon graduation. Remember from Chapter 1 that there are many career paths in psychology. In 2019, over 1.5 million individuals with bachelor's degrees in psychology were employed in many work sectors.

Source: Cartoon Resource/ Shutterstock.com.

Archival Research

Archives are also valuable sources of historical information about psychology itself. Through the efforts of John Popplestone and Marion McPherson (L. T. Benjamin, 2002), the Archives of the History of American Psychology at The University of Akron has become the main repository of records providing insight into the major issues, pioneers, and landmark events in the history of American psychology (Popplestone & McPherson, 1976). One of your textbook authors visited this fascinating museum (though perhaps not where you want to go for spring break!).

Source: Goritza/Shutterstock.com; Triff/Shutterstock.com; B.erne/Shutterstock.com; Fer Gregory/Shutterstock.com.

Beck Depression Inventory-II (Canel-Çınarbaş et al., 2011), and the Korean version of the Panic Disorder Severity Scale (E. H. Lee et al., 2009). However, it is important to ensure that the research materials, samples, and participants' familiarity with the research task are culturally equivalent, or appropriate for the cultures under study (J. Allen & Walsh, 2000).

Archival Research

archival research The systematic examination of collections of letters, manuscripts, recordings, or other records.

The largest potential source of knowledge from descriptive research is **archival research**, which examines collections of letters, manuscripts, recordings, or similar materials. The uses of archival research are virtually without limit. For example, archival research discussed in Chapter 3 attempts to answer the following controversial question: Do right-handed people live longer than left-handed people? An archival study of North American comic books found that their number of authoritarian themes increased during times of high perceived national social and economic threat and decreased during times of low perceived national social and economic threat (B. E. Peterson & Gerstein, 2005). And consider this question: Do people of different genders have different or similar physical fitness goals? In one archival study, researchers conducted Google Image searches for four consecutive years for the terms "burn fat" and "build muscle." They also recorded the likely gender of the person who appeared in the image. The researchers found that, regardless of the year, images of women were associated with the term "burn fat" and images of men were associated with the term "build muscle." This finding indicates that there are indeed gender differences in physical fitness goals (Salvatore & Maracek, 2010).

Archival research can be descriptive or, as in the gender exercise study described previously, it can be correlational (discussed next). But note that, as is true of all descriptive and correlational research, archival research does not permit conclusive causal statements about its findings. For example, archival research cannot determine why women are more interested in burning fat than building muscle. Nor can archival research, by itself, explain why comic book themes become more authoritarian during times of perceived national social and threat.

correlational research Research that studies the degree of relationship between two or more variables.

correlation The degree of relationship between two or more variables.

variable An event, behavior, condition, or characteristic that has two or more values.

Correlational Research

When psychologists want to predict changes in one variable based on changes in another, rather than simply describe something, they turn to **correlational research**. A **correlation** refers to the degree of relationship between two or more *variables*. A **variable** is an event, behavior, condition, or characteristic that has two or more values. Examples of possible variables include age, height, temperature, and intelligence.

Kinds of Correlation

A **positive correlation** between two variables indicates that they tend to change values in the same direction. That is, as the first increases, the second increases, and as the first decreases, the second decreases. For example, as hours of studying increase, grade-point average (GPA) tends to increase. A **negative correlation** between two variables indicates that they tend to change values in opposite directions. For example, as age increases in adulthood, visual acuity tends to decrease. Correlations range in magnitude from zero, meaning that there is no systematic relationship between the two variables, to 1.00, meaning that there is a perfect relationship between them. Thus, a perfect positive correlation would be +1.00, and a perfect negative correlation would be –1.00.

positive correlation A correlation in which variables tend to change values in the same direction.

negative correlation A correlation in which variables tend to change values in opposite directions.

Consider the relationship between obesity and exercise. The more people exercise, the less they tend to weigh. This relationship indicates a negative correlation between exercise and body weight: As one increases, the other decreases. When two variables are correlated, one can be used to *predict* the other, but the first does not necessarily have a causal relationship with the other (T. A. Brigham, 1989). That is, *correlation* does not necessarily imply **causation**. You will hear that sentence again and again throughout your study of psychology. Even though it is plausible that exercise causes lower body weight, it is also possible that the opposite is true: Lower body weight might cause people to exercise. Lighter people might exercise more because they find it less strenuous, less painful, and less embarrassing than heavier people do. Nonetheless, correlational research plays an important role when experimental research is either unethical or impractical to conduct, such as in educational settings (B. Thompson et al., 2005) or in the study of mental health disorders (e.g., you cannot randomly assign some people to get depressed).

causation An effect of one or more variables on another variable.

Causation Versus Correlation

As another example of the need to distinguish between causation and correlation, consider the positive correlation between socioeconomic status and the likelihood of developing a deadly form of skin cancer called malignant melanoma (Johnson-Obaseki et al., 2015). This positive correlation means that as income level rises, the probability of getting the disease also rises. You would be correct in predicting that people who attend college will be more likely, later in life, to develop malignant melanoma than will people who never go beyond high school.

But does this finding mean that you should drop out of school today to avoid the disease? The answer is no, because the positive correlation between income level and malignant melanoma does not necessarily mean that attending college causes the disease. Other factors common to people who attend college might cause them to develop the disease. Given that extensive exposure to the sun is a risk factor in malignant melanoma (Ivry et al., 2006), perhaps people who attend college increase their risk of malignant melanoma by exposing themselves to the sun more frequently than do those who have only a high school education. College students might be more likely to spend spring breaks in Florida (not in the History of Psychology museum!), find summer jobs at beach resorts, or go on frequent Caribbean vacations after graduating, and finding higher-paying, full-time jobs. Instead of dropping out of college to avoid the disease, students might be wiser to spend less time in the sun and to make regular use of a highly protective sunscreen.

Causation Versus Correlation

It turns out that there is a strong positive correlation between ice cream sales and violence. When ice cream sales rise, so do homicides. Of course, the creamy cones do not actually cause violence (except when they are sold out of your favorite flavor!). There is a third variable—hot weather—that causes both an increase in homicides and ice cream consumption. In fact, every 9-degree Fahrenheit rise in temperature translates to about a 9% increase in homicides over the next week in New York and Chicago (Xu et al., 2020).

Source: Asier Romero/Shutterstock.com.

Psychologists are careful not to confuse causation and correlation. If two variables are positively correlated, the first might cause changes in the second, the second might cause changes in the first, or another variable might cause changes in both. Because of the difficulty in distinguishing causal relationships from mere correlational ones, correlational research has stimulated controversies in important areas of research. Does televised violence cause real-life aggression? A review of research on that question found a significant positive correlation between watching televised violence and exhibiting aggressive behavior. But this correlation does not indicate that televised violence *causes* aggressive behavior (J. L. Freedman, 1984). Perhaps people who are aggressive for other reasons simply prefer to watch violent television programs. Nonetheless, as discussed in

Chapters 7 and 17, there have been a number of experimental studies that do support a causal link between media violence and viewer aggression (Bushman & Anderson, 2001).

Experimental Research

The research methods discussed so far do not enable us to discover causal relationships between variables. Even when there is a strong correlation between variables, we cannot presume a causal relationship between them. To determine whether there is a causal relationship between variables, scientists use the **experimental method**. Psychologists have relied on the experimental method ever since the discipline finally completed its separation from philosophy in the late 19th century (Hatfield, 2002).

experimental method Research that manipulates one or more variables, while controlling other factors, to determine the effects on one or more other variables.

independent variable A variable manipulated by the experimenter to determine its effect on another, dependent, variable.

dependent variable A variable showing the effect of the independent variable.

experimental group Participants in an experiment who are exposed to the experimental condition of interest.

control group The participants in an experiment who are not exposed to the experimental condition of interest.

field experiment An experiment that is conducted in real life as opposed to laboratory settings.

placebo An inactive substance that might induce some of the effects of the drug for which it has been substituted.

Experimental Method

Every experiment includes at least one *independent variable* and one *dependent variable*. The **independent variable** is manipulated by experimenters, which means that they determine its values before the experiment begins. The **dependent variable** is what the experimenters measure to determine any effects of the independent variable. In terms of cause-and-effect relationships, the independent variable would be the *cause* and changes in the dependent variable would be the *effect*. Thus, in an experiment on the effects of drinking on driving, the independent variable of alcohol intake would be the cause of changes in the dependent variable of accident frequency on the driving simulator. See Chapter 6 for a discussion of some fascinating experiments from David Strayer's driving lab at the University of Utah.

The simplest experiment uses one independent variable with two values (an experimental condition and a control condition) and one dependent variable. A group of participants, the **experimental group**, is exposed to the experimental condition, and a second group of participants, the **control group**, is exposed to the control condition. The control condition is often simply the absence of the experimental condition. For example, the experimental condition might be exposure to a particular advertisement, and the control condition might be no exposure to the advertisement. The dependent variable might be the number of sales of the advertised product. The control group provides a standard of comparison for the experimental group. If you failed to include a control group in the suggested experiment on the effects of advertising, you would be unable to determine whether the advertising accounted for changes in the volume of sales. This example illustrates a **field experiment**, which is conducted in real life, as opposed to laboratory settings.

To appreciate the nature of the experimental method, imagine you are a psychologist interested in conducting an experiment on the effect of melatonin on nightly sleep duration. A basic experiment on this topic is illustrated in Table 2-2. Assume that introductory psychology students volunteer to be participants in the study. Members of the experimental group receive the same dose of melatonin nightly for 10 weeks, whereas members of the control group receive a **placebo**, which has no demonstrated effect on sleep (but the participants do not know that the pill is inert). As the experimenter, you would try to keep constant all other factors that might affect the two groups. By treating both groups the same except for the condition to which the experimental group is exposed, you would be able to conclude that any significant difference in average sleep duration between the experimental group and the control group was probably caused by the experimental group's receiving doses of melatonin. Without the use of a control group, you would have no standard of comparison and would be less secure in reaching that conclusion.

In the experiment on melatonin and nightly sleep duration, the independent variable (drug condition) has two values: melatonin and placebo. The experimenter is interested in the effect of the independent variable on the dependent variable. The dependent

"On your mark!"

Experimental Variables

In this cartoon, the hypothetical independent variable could be anything from sex differences to alcohol intake in the two rats and the dependent variable would be the time taken to complete the maze.

Source: Cartoon Resource/Shutterstock.com.

TABLE 2-2 A Basic Experimental Research Design

Group	Independent Variable	Dependent Variable
(Randomly assigned)	(Drug condition)	(Sleep)
Experimental group	Take melatonin	Hours of sleep
Control group	Take placebo	Hours of sleep

variable in this case is nightly sleep duration, with many possible values: 6 hours and 2 minutes, 7 hours and 48 minutes, and so on.

The effectiveness of melatonin in treating insomnia was, in fact, supported by a double-blind, placebo study of children with insomnia. Forty children, ages 6 to 12 years, with chronic difficulty falling asleep were randomly assigned to groups that received doses of melatonin or a placebo at 6 p.m. for four weeks. Thus, drug condition was the independent variable. The dependent variables included the time when the children turned the lights off in their bedrooms, the time when they fell asleep, and how long they slept. Those who received a placebo showed no significant change on these variables. In contrast, those who received doses of melatonin turned their lights off an average of 34 minutes earlier, fell asleep an average of 75 minutes earlier, and slept an average of 41 minutes longer (Smits et al., 2001). Placebo control groups are essential in research on drug therapy, as in research on the effectiveness of medication for treating depression (J. Hughes et al., 2012) or anxiety (Bidzan et al., 2012).

Internal Validity

An experimenter must do more than simply manipulate an independent variable and record changes in the dependent variable. The experimenter must also ensure the **internal validity** of the experiment by *controlling* any extraneous factors whose effects on the dependent variable might be confused with those of the independent variable (Christ, 2007) in order to show a cause-and-effect relationship. Such extraneous factors are called **confounding variables**, because their effects are confused, or *confounded*, with those of the independent variable. A confounding variable might be associated with the experimental situation, participants, or experimenters involved in an experiment.

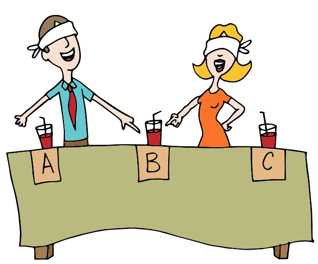
internal validity The extent to which changes in a dependent variable can be attributed to one or more independent variables rather than to a confounding variable.

confounding variable A variable whose unwanted effect on the dependent variable might be confused with that of the independent variable.

Situational Variables In carrying out the procedure in the melatonin experiment, you would not want any confounding variables to affect nightly sleep duration. You would want the participants to be treated the same, except that those in the experimental group would receive their dose of melatonin nightly over a 10-week period. But suppose that some participants in the experimental group decided to take other sleeping pills, to exercise more, or to practice meditation. If, at the end of the study, the experimental group had a longer nightly sleep duration than the control group, the results might be attributable not to the melatonin but to confounding variables—that is, differences between the groups in the extent to which they employed those other tools not relevant to your experiment.

As an example of the importance of controlling potential confounding procedural variables, consider what happened when the Pepsi-Cola company conducted a "Pepsi Challenge" taste test, an example of *consumer psychology* ("Coke-Pepsi Slugfest," 1976). Coca-Cola drinkers were asked to taste each of two unidentified cola drinks and state their preference. The drinks were Coca-Cola and Pepsi-Cola. The brand of cola was the independent variable, and their preference was the dependent variable. To keep the participants from knowing which cola they were tasting, they were given Pepsi in a cup labeled M and Coke in a cup labeled Q. To the delight of Pepsi Co stockholders, most of the participants preferred Pepsi.

The Pepsi-Cola company proudly—and loudly—advertised this finding as evidence that even Coke drinkers preferred Pepsi. But knowing the pitfalls of experimentation, the Coca-Cola company replicated the experiment, this time filling both cups with Coca-Cola.



Consumer Psychology: Example of a Blind Taste Test

How could you control for the effect of labeling each of the three products?

Source: John T Takai/Shutterstock.com.

Most of the participants still preferred the cola in the cup labeled M. Evidently, the Pepsi Challenge had demonstrated only that Coca-Cola drinkers preferred the letter M to the letter Q. The effect of the letters on the dependent variable (the taste preference) had been confounded with that of the independent variable (the kind of cola).

If you were asked to design a more scientifically sound Coke-Pepsi taste challenge, how would you control the effect of the letter on the cup? Pause to think about this question before reading on. One way to control it would be to use cups without letters. Of course, the experimenter would have to keep track of which cup contained Coke and which contained Pepsi. A second way to control the effect of the letter would be to label each of the colas M on half of the taste trials and Q on the other half. Thus, two ways to control potential confounding procedural variables are to eliminate them or to ensure that they affect all conditions equally.

Participant Variables Experimenters must likewise control potential confounding participant variables that might produce effects that would be confused with those of the independent variable. Suppose that in the melatonin experiment the participants in the experimental group initially differed from the participants in the control group on their nightly sleep duration before the study even began. These differences might carry over into the experiment, affecting the participants' nightly sleep duration during the course of the study and giving the false impression that the independent variable (melatonin versus no melatonin) caused a significant difference on the dependent variable (nightly sleep duration) between the two groups.

Experimenters increase the chance that the experimental group and the control group will be initially equivalent on as many participant variables as possible by relying on *random assignment* of participants to groups (Enders et al., 2006). In **random assignment**, participants are as likely to be assigned to one group as to another. Given a sufficiently large number of participants, random assignment will make the two groups initially equivalent on many, though not necessarily all, relevant participant variables.

After randomly assigning participants to the experimental group and the control group, you still would have to control for other participant variables. One of the most important of these is **participant bias**, the tendency of people who know they are participants in a study to behave differently than they normally do. As in the case of naturalistic observation, you might choose to be unobtrusive, exposing people to the experimental condition without their being aware of it. If this were impossible, you might choose to misinform the participants about the true purpose of the study. (The ethical issues involved in using deception are discussed in the section entitled "Ethics of Psychological Research.") Placebos are used in conjunction with random assignment so that participants do not succumb to demand characteristics of the experimental situation, in which knowledge of the experimental hypothesis leads them to perform in a manner that supports it—even more so when they like the experimenter (A. L. Nichols & Maner, 2008).

Experimenter Variables Experimenters must control not only potential confounding variables associated with the research procedure or the research participants but also potential confounding variables associated with themselves. *Experimenter effects* on dependent variables can be caused by the experimenter's personal qualities, actions, and treatment of data. Experimenter effects have been studied most extensively by Robert Rosenthal and his colleagues, who have demonstrated them in many studies since the early 1960s. The experimenter's personal qualities—including gender, attire, and attractiveness—can affect participants' behavior (M. L. Barnes & Rosenthal, 1985).

Also of concern is the effect of the experimenter's actions on the recording of data or on the participants' behavior, as in the **experimenter bias effect**. This occurs when the results are affected by the experimenter's expectancy about the outcome of a study, which is expressed through their unintentional actions. The tendency of participants to behave in accordance with experimenter expectancy is called *self-fulfilling prophecy*. Actions that might promote self-fulfilling prophecy include facial expressions (perhaps smiling at participants in one group and frowning at those in another), mannerisms (e.g., shaking hands with participants in one group but not with those in another), or tone of voice (such

random assignment The assignment of participants to experimental and control conditions so that each participant is as likely to be assigned to one condition as to another.

participant bias The tendency of people who know they are participants in a study to behave differently than they normally would.

experimenter bias effect The tendency of experimenters to let their expectancies alter the way they treat their participants.

Self-Fulfilling Prophecy

Can teacher expectations affect student performance? See Chapter 17 on Social Psychology.

Source: Andresr/Shutterstock.com.

The Research Process

Can Experimenter Expectancies Affect the Behavior of Laboratory Rats?

Rationale

Robert Rosenthal noted that, in the early 20th century, Ivan Pavlov had found that each succeeding generation of his animal subjects learned tasks faster than the preceding one. At first, Pavlov presumed that this improvement supported the (since-discredited) notion of the inheritance of acquired characteristics. But he eventually came to believe that the animals' improvement was caused by changes in the way in which his experimenters treated them. Rosenthal decided to determine whether experimenter expectancies could likewise affect the performance of laboratory animals.

Source: sextoacto/Shutterstock.com.

Method

Rosenthal and his colleague Kermit Fode had 12 students act as experimenters in a study of maze learning in rats conducted at Harvard University (Rosenthal & Fode, 1963). Six of the students were told that their rats were specially bred to be "maze bright," and six were told that their rats were specially bred to be "maze dull." In reality, the rats did not differ in their inborn maze-learning potential. Each student was given five rats to run in a T-shaped maze. The rats received a food reward on alternating arms and the researched controlled for side bias. In other words the rats learned to respond to the baited arm rather than to the direction left or right. The students ran the rats 10 times a day for five days and recorded how long it took them to reach the food.

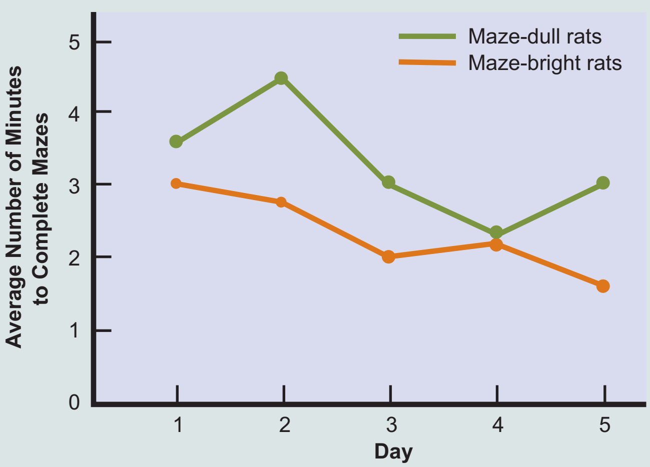


FIGURE 2-1 Experimenter Bias

The graph shows the mean results of the Rosenthal and Fode (1963) experiment, which found that allegedly maze-bright rats ran mazes faster than allegedly maze-dull rats did, in accordance with the experimenters' expectations.

Source: Rosenthal & Fode (1963).

Results and Discussion

As shown in Figure 2-1, the results indicated the apparent influence of experimenter expectancy: On average, the "maze-bright" rats ran mazes faster than the "maze-dull" rats did. Because there was no evidence of cheating or misrecording of data by the students, the researchers attributed the results to experimenter expectancy. The students' expectancies apparently influenced the manner in which they trained or handled the rats, somehow leading the rats to perform in accordance with their predictions. For example, those who trained "maze-bright" rats reported handling them more, and more gently, than did those who trained "maze-dull" rats. Confidence in the experimenter expectancy effect with animal subjects was supported in a replication study by a different researcher, using different rats, and involving a different learning task (Elkins, 1987). These findings indicate that those responsible for handling animals during an experiment should, if possible, be kept unaware of any presumed differences among the animals or even of the experimental hypothesis.

as speaking in an animated voice to participants in one group and speaking in a monotone to those in another). Self-fulfilling prophecy is especially important to control in studies of psychotherapy, because therapist expectancies, rather than the therapy itself, might affect the outcome of treatment (M. J. Harris, 1994).

The classic research study discussed in "The Research Process" box demonstrated that experimenter expectancies can even affect the behavior of animal subjects.

Gold Standard for Clinical Trials

Randomized controlled trials are study designs that randomly assign participants into an experimental or control group. Double-blind studies are the gold standard because they result in high levels of evidence without bias. Triple-blind studies are randomized; the treatment or intervention assignment is unknown to the research participant and the experimenter, and group assignment is also not known to the individual who assesses the outcomes (such as the statistician).

Source: bangoland/Shutterstock.com.

How might experimenter bias affect the results of the melatonin experiment? The experimenter might inadvertently interact more with or give sleep hygiene advice to the participants in the experimental group, perhaps causing them to sleep better than they would have otherwise. One way to control for experimenter bias is to make sure that people who interact with the participants are unaware of the research hypothesis, eliminating the influence of the experimenter's expectancies on the participants' performance.

double-blind technique A procedure that controls experimenter bias and participant bias by preventing experimenters and participants from knowing which participants have been assigned to particular conditions.

At times, both participant bias and experimenter bias might become confounding variables. This possibility might prompt experimenters to use the **double-blind technique** or *double-masking*, in which neither the experimenter nor the participants know the conditions to which the participants have been assigned. This technique is common in studies of the effectiveness of drug treatments for psychological disorders. Consider a study on the effectiveness of nicotine-replacement therapy for smoking cessation. In a double-blind study, experimental groups received either nicotine gum or a nicotine inhaler, while the control group received a placebo. Participants in the nicotine replacement condition were significantly more likely to smoke fewer cigarettes or stop smoking altogether than the participants in the placebo condition (Kralikova et al., 2009). In a double-blind melatonin experiment described above, one group gets melatonin and the other a placebo (a sugar pill that looks like it could be melatonin). Neither the experimenter nor the participants would know which participants received the melatonin and which received the placebo. And any results (changes in the dependent variable of sleep duration or quality) would be caused by the melatonin.

External Validity

external validity The extent to which the results of a research study can be generalized to other people, animals, or settings.

Though experimenters are chiefly concerned with matters of internal validity, they are also ultimately interested in **external validity**—the extent to which they can generalize their research findings to other populations, settings, and procedures. Experiments may have strong internal validity but lack external validity. Researchers in the field of alcoholism treatment, for example, note that treatment studies with excellent internal validity often bear little relationship to what is actually done in normal clinical practice, thus potentially limiting their external validity (Sterling, 2002). Similarly, laboratory experiments on consumer behavior may have strong internal but weak external validity in regard to consumer behavior in everyday life, making those who run marketing campaigns nervous (Winer, 1999). For instance, even when the internal validity of the Coke/Pepsi experiment described previously was improved, the study still lacked external validity: Some participants preferred their sip of Pepsi simply because it was sweeter, but, in real life, most people drink the whole can/bottle rather than merely taking a sip, and the initial sweetness might not be valued as highly after 50 sips (Gladwell, 2005). Likewise, the external validity of experiments on driving behavior using laboratory simulators depends on the simulation being relevant to real-world driving situations (Araújo, 2007).

Moreover, as stressed by Stanley Sue (1999) and other psychologists who highlight the sociocultural perspective, the results of a research study done in one culture will not necessarily be generalizable to another culture or ethnic group. Researchers must identify the specific populations to which their research findings may be applied. An archival study of 14 psychology journals over a five-year period found that only 61% of 2,536 articles related to applied psychology reported participants' ethnicity. Of those that did, the ethnic breakdown of the samples was generally representative of the population estimates provided by the U.S. Census Bureau. However, Latinx people—for whom English may be a second language—were underrepresented. The generalizability of research findings from those studies to non-English speakers may thus be limited (Case & Smith, 2000). Overall, a 2008 survey of the top psychology journals found that 96% of research participants were from Western industrialized countries—which house a mere 12% of the world's population. The fact that most classic psychology studies were done with people from Western, educated, industrialized, rich, and democratic **(WEIRD)** societies—and particularly American undergraduates—presents a challenge to the external validity of this research (Henrich et al., 2010). The narrow focus of WEIRD research can create bias sampling and deserves attention and consideration when using a cultural lens for interpretation.

WEIRD An acronym for Western, Educated, Industrialized, Rich, and Democratic samples.

But cross-cultural replications of research can and do demonstrate the possible universality of psychology findings. Research has found cross-cultural similarities in coping strategies employed by Canadian and New Zealand women with a history of child sexual abuse (Barker-Collo et al., 2012). The same sociocultural factors that were associated with body self-image and eating disorders in American women also were found in Japanese women (Yamamiya et al., 2008). And researchers have found that marital discord also is associated with depression in both American and Brazilian women (Hollist et al., 2007). Furthermore, CBT works worldwide to treat psychological disorders albeit with some specific modifications to adapt it for different cultures (Naeem, 2019).

Replication to assess external validity is not only important in experimental research but in other kinds of research as well. An American survey found that about the same percentage of people had obsessions or compulsions as several previous national surveys had found (Ruscio et al., 2010). A study that used psychological testing of self-esteem replicated prior research by showing that a major factor in overall self-esteem is one's perceived appraisal by significant others such as parents, teachers, and friends (Y. Stephan & Maiano, 2007). And a study on the effectiveness of a high school suicide prevention program found results similar to that of a previous study in regard to changing undesirable attitudes toward suicide and in decreasing reluctance to seek mental health treatment (Ciffone, 2007).

Another potential problem that may affect external validity is the use of volunteer participants. People who voluntarily take part in a given experiment might differ from people who do not, like the presidential polling described earlier, thereby limiting the generalizability of the research findings. In a study using volunteer participants, undergraduates were given the choice of participating in either a study in which they would be given a sexual interview or a study in which they would watch an explicit sexual video. When compared with students who refused to volunteer, students who volunteered for either of the studies were more sexually experienced, held less traditional sexual attitudes, and scored higher on measures of sexual esteem and sexual sensation seeking. These findings indicate that people who participate in sexual research might not be representative of people in general, limiting the confidence with which sex researchers can generalize their findings (Wiederman, 1999b). Would you watch pornography in the name of science?

Of course, differences between volunteers and nonvolunteers—just like cultural differences—do not automatically mean that the results lack external validity. The best way to determine whether the research has external validity is to replicate it. Most replications are approximate; they rarely use the same setting, participants, or procedures. The ideal would be to replicate studies systematically several times, varying one aspect of the study each time (Hendrick, 1990). Thus, you would be more confident in your ability to generalize the findings of the melatonin experiment if people with insomnia, of a variety of ages, in several different cultures, succeeded in sleeping longer after habitually taking melatonin before bedtime.

Now that you have been introduced to the descriptive, correlational, and experimental methods of research, you should be able to recognize them as you read about research studies described in later chapters. As you read particular studies, try to determine which kind of method was used as well as its possible strengths and weaknesses—most notably, any potential confounding variables and any limitations on the generalizability (external validity) of the research findings. You are now ready to learn about how psychologists analyze the data generated by these research methods.

Section Review: Methods of Psychological Research

1. Why is it important to use unbiased samples in surveys?
2. What is validity in psychological testing?
3. What is an independent variable?
4. What is external vs. internal validity?

Statistical Analysis of Research Data

How would you make sense out of the data generated by the melatonin experiment? In analyzing the data, you would have to do more than simply state that Ann Begay slept 9.1 hours, Steve White slept 7.8 hours, Sally Ramirez slept 8.2 hours, and so on. You would have to identify overall patterns in the data and determine whether the data support the research hypothesis that inspired the experiment.

As mentioned earlier, to make sense out of data, psychologists rely on statistics. The term *statistics* was originally used to refer to the practice of recording quantitative political and economic information about European nation-states (Cowles, 1989). During the 20th century, the use of statistics to analyze research data became increasingly more prevalent in articles published in psychology journals (S. Parker, 1990). Psychologists use *descriptive statistics* to summarize data, *correlational statistics* to determine relationships between variables, and *inferential statistics* to test their experimental research hypotheses. Appendix C (available in the Online Edition) presents an expanded discussion of statistics and their calculation, and all psychology majors will take a full course in this at some point in their studies.

Descriptive Statistics

descriptive statistics Statistics that summarize research data.

You summarize your data by using **descriptive statistics**. An early champion of the use of descriptive statistics was Florence Nightingale (1820–1910), one of the founders of modern nursing. She urged that hospitals keep medical records on their patients and demonstrated statistically that British soldiers during times of war were more likely to die from disease and unsanitary conditions than from combat. She also was a pioneer in the use of graphs to support her conclusions. Her work led to reforms in nursing and medicine and to her being made a fellow of the Royal Statistical Society and an honorary member of the American Statistical Association (Viney, 1993). Descriptive statistics include *measures of central tendency* and *measures of variability*.

measure of central tendency A statistic that represents the "typical" score in a set of scores.

mode The score that occurs most frequently in a set of scores.

median The middle score in a set of scores that have been ordered from lowest to highest.

Measures of Central Tendency

A **measure of central tendency** is a single number used to represent a set of scores. The measures of central tendency include the *mode*, the *median*, and the *mean*. Psychological research uses the mode least often, the median somewhat more often, and the mean most often.

The **mode** is the most frequent score in a set of scores. As shown in Table 2-3, in the melatonin experiment the mode for the experimental group is 8.6 hours and the mode for the control group is 8.9 hours. The **median** is the middle score in a set of scores that have

TABLE 2-3 Descriptive Statistics From a Hypothetical Experiment on the Effect of Melatonin on Average Nightly Sleep Duration

Experimental Group (Melatonin)				Control Group (No Melatonin)			
Participant	Duration	*d*	d^2	Participant	Duration	*d*	d^2
1	9.1	0.2	0.04	1	7.4	–0.5	0.25
2	8.6	–0.3	0.09	2	8.2	0.3	0.09
3	8.6	–0.3	0.09	3	9.5	1.6	2.56
4	8.8	–0.1	0.01	4	8.9	1.0	1.00
5	7.8	–1.1	1.21	5	6.7	–1.2	1.44
6	9.9	1.0	1.00	6	8.9	1.0	1.00
7	8.6	–0.3	0.09	7	7.5	–0.4	0.16
8	9.7	0.8	0.64	8	6.2	–1.7	2.89
9	9.0	0.1	0.01	9	7.8	–0.1	0.01
	Sum = 80.1		Sum = 3.18		Sum = 71.1		Sum = 9.40

Experimental Group (Melatonin):

Mode = 8.6 hours

Median = 8.8 hours

Mean = $\frac{80.1}{9}$ = 8.9 hours

Range = 9.9 – 7.8 = 2.1 hours

Variance = $\frac{\text{sum of } d^2}{\text{no. of participants}} = \frac{3.18}{9} = 0.35$

Standard deviation $= \sqrt{\text{Variance}}$

$= \sqrt{0.35}$

= 0.59 hours

Control Group (No Melatonin):

Mode = 8.9 hours

Median = 7.8 hours

Mean = $\frac{71.1}{9}$ = 7.9 hours

Range = 9.5 – 6.2 = 3.3 hours

Variance = $\frac{\text{sum of } d^2}{\text{no. of participants}} = \frac{9.40}{9} = 1.04$

Standard deviation $= \sqrt{\text{Variance}}$

$= \sqrt{1.04}$

= 1.02 hours

Note: *d* = deviation from the mean.

been arranged in numerical order. Thus, in the melatonin experiment the median score for each group is the fifth score after the scores are put in rank order. The median for the experimental group is 8.8 hours and the median for the control group is 7.8 hours. You are most familiar with the **mean**, which is the arithmetic average of a set of scores. You use the mean when you calculate your exam average, batting average, or the average gas mileage of your car. In the melatonin experiment, the mean for the experimental group is 8.9 hours and the mean for the control group is 7.9 hours.

mean The arithmetic average of a set of scores.

One of the problems in the use of measures of central tendency is that they can be used selectively to create misleading impressions. Suppose you had the following psychology exam scores: 23, 23, 67, 68, 69, 70, 91. The mode (the most frequent score) would be 23, the median (the middle score) would be 68, and the mean (the average score) would be 58.7. In this case, you would prefer the median as representative of your performance. But what if you had the following scores: 23, 67, 68, 69, 70, 91, 91? The mode would be 91, the median would be 69, and the mean would be 68.43. In that case, you would prefer the mode as representative of your performance.

Product advertisers, government agencies, and political parties also are prone to this selective use of measures of central tendency, as well as other statistics, to support their claims.

Measures of Variability

measure of variability A statistic describing the degree of dispersion in a set of scores.

range A statistic representing the difference between the highest and lowest scores in a set of scores.

To represent a distribution of scores, psychologists do more than report a measure of central tendency. They also report a **measure of variability**, which describes the degree of dispersion of the scores or how spread out they are. That is, do the scores tend to bunch together, or are they scattered? Commonly used measures of variability include the range and the standard deviation. The **range** is the difference between the highest and the lowest score in a set of scores. In Table 2-3, the range of the experimental group is 9.9 – 7.8 = 2.1 hours, and the range of the control group is 9.5 – 6.2 = 3.3 hours. But the range can be misleading because one extreme score can create a false impression. Suppose that a friend conducts a similar experiment and reports that the range of sleep duration among the 15 participants in his experimental group is 4 hours, with the longest duration being 9.3 hours and the shortest duration being 5.3 hours. You might conclude that there was a great deal of variability in the distribution of scores. But what if he then reported that only one participant slept less than 9.1 hours? Obviously, the scores would bunch together at the high end, making the variability of scores much less than you had presumed.

standard deviation A statistic representing the degree of dispersion of a set of scores around their mean.

variance A measure based on the average deviation of a set of scores from their group mean.

Because of their need to employ more meaningful measures of variability than the range, psychologists prefer to use the standard deviation. The **standard deviation** represents the degree of dispersion of scores around their mean and is the square root of a measure of variability called the *variance*. The **variance** is a measure based on the average deviation of a set of scores from their group mean. Table 2-3 shows that the standard deviation of the experimental group is 0.59 hours, whereas the standard deviation of the control group is 1.02 hours. Thus, the distribution of scores in the experimental group has a larger mean, but the distribution of scores in the control group has a larger standard deviation, meaning they are more spread out.

Correlational Statistics

coefficient of correlation A statistic that assesses the degree of association between two or more variables.

If you were interested in predicting one set of scores from another, you would use a measure of correlation. The concept of correlation was put forth in 1888 by Francis Galton, who wanted a way to represent the relationship between parents and offspring on factors, such as intelligence, presumed to be affected by heredity. Whereas the mean and standard deviation are useful in describing individual sets of scores, a statistic called the coefficient of correlation is useful in quantifying the degree of association between two or more sets of scores (variables). The **coefficient of correlation** was devised by the English mathematician Karl Pearson (1857–1936) and is often called *Pearson's r* (with the *r* standing for "regression," another name for correlation). As you learned earlier, a correlation can be positive or negative and can range from zero to +1.00 or –1.00. The types of correlations are illustrated graphically in Figure 2-2.

Inferential Statistics

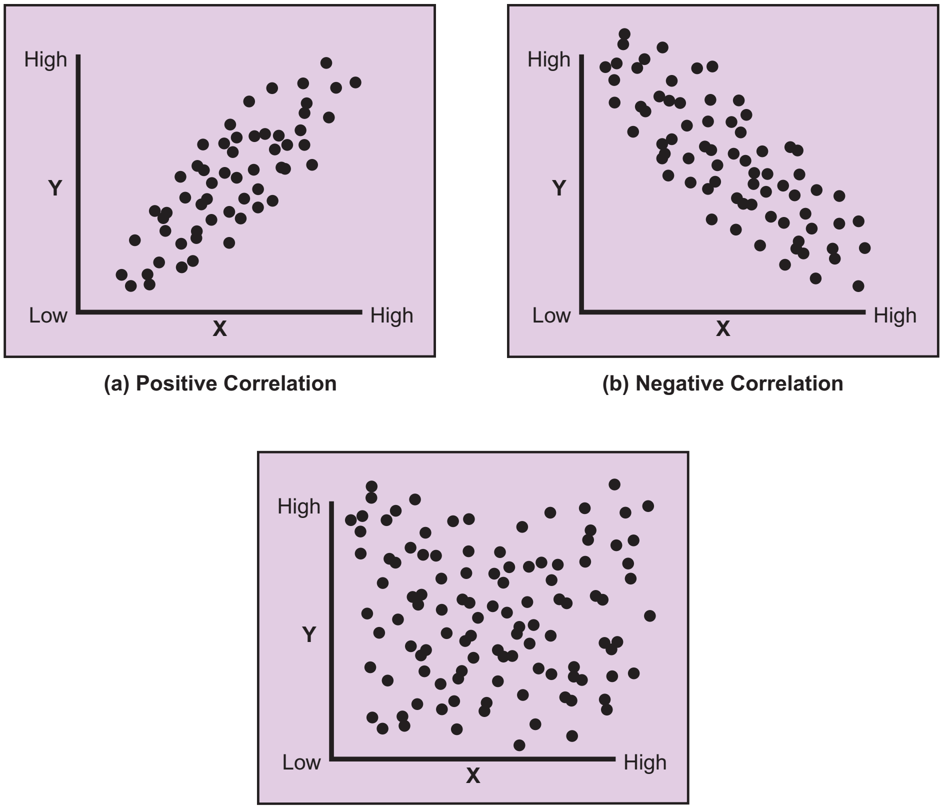
inferential statistics Statistics used to determine whether changes in a dependent variable are caused by an independent variable.

In the melatonin experiment, the experimental group had a longer average nightly sleep duration than the control group. But is the difference in average nightly sleep duration between the two groups large enough to conclude with confidence that melatonin was responsible for the difference? Perhaps the difference happened by chance—that is, because of a host of random factors unrelated to melatonin. To determine whether the independent variable, rather than chance factors, caused the changes in the dependent variable, psychologists use **inferential statistics**. By permitting psychologists to determine the causes of events, inferential statistics help them achieve the goal of explanation and make inferences from the samples used in their experiment to the populations of individuals they represent.

statistical significance A low probability (usually less than 5%) that the results of a research study are due to chance factors rather than to the independent variable.

Statistical Significance

If there is a low probability that the difference between groups on the dependent variable is attributable to chance (that is, to random factors), the difference is statistically significant and is attributed to the independent variable. The concept of **statistical significance**



(a) Positive Correlation **(b) Negative Correlation** **(c) Zero Correlation**

FIGURE 2-2 Correlations
(a) In a *positive correlation*, scores on the variables increase and decrease together. An example is the relationship between SAT verbal scores and college GPA. (b) In a *negative correlation*, scores on one variable increase as scores on another variable decrease. An example is the relationship between age and nightly sleep. (c) In a *zero correlation*, scores on one variable are unrelated to scores on another. An example is the relationship between the amount of tea you drink and your intelligence (despite what your tea-loving friends might tell you!).

was first put forth by the English mathematician Ronald Fisher (1890–1962) when he sought a way to test a noblewoman's claim that she could tell whether tea or milk had been added to her cup first (Tankard, 1984). Though he never carried out the demonstration, he proposed presenting her with a series of cups in which tea was sometimes added first and milk was sometimes added first. He assumed that if she could report the correct order at a much greater than chance level, her claim would be verified. To rule out simple lucky guessing, she would have to be correct significantly more than 50% of the time—the chance level of guessing between two events.

In the melatonin experiment described previously, you would expect that chance factors would account for some changes in the sleep duration of participants in both groups during the course of the study. As a result, for the difference in average sleep duration between the two groups to be statistically significant, it would have to be significantly larger than would be expected by chance alone. Psychologists usually accept a difference as statistically significant when there is less than a 5% (5 in 100) probability that the difference is the product of chance factors—the so-called .05 alpha level.

Nonetheless, even when the analysis of research data reveals statistical significance, the best way to determine whether research findings are real and generalizable is—you guessed it—to replicate them (R. Falk, 1998). Two actual experiments did, in fact, "replicate" the findings of the imaginary melatonin experiment described earlier. These experiments, which used the double-blind technique, found that participants who took melatonin slept longer than did participants who took the placebo—regardless of whether the participants were normal sleepers (Waldhauser et al., 1990) or insomnia sufferers (MacFarlane et al., 1991). A meta-analysis (discussed next) concluded that "although the absolute benefit of melatonin compared to placebo is smaller than other pharmacological treatments for insomnia, melatonin may have a role in the treatment of insomnia given its relatively benign side-effect profile compared to these agents" (Ferracioli-Oda et al., 2013, p. e63773). Thus, there is scientific support for the claims made in the *Newsweek* cover story that opened this chapter.

Meta-analysis

meta-analysis A technique that combines the results of many similar studies to determine the effect size of a particular kind of independent variable.

Still another approach to assessing generalizability is to use the relatively new statistical technique called **meta-analysis**. Meta-analysis combines research findings from many, typically dozens, of related studies and goes beyond simply determining statistical significance. After gathering the studies under analysis, the researcher computes a statistic called an *effect size* (e.g., *Cohen's d*) for each study in the analysis. Then, the effect size statistic is averaged across all studies (often weighted so that larger or better studies count for more) to compute the average size of the effect of the independent variable. As a general rule, effect sizes are described as small ($d = .20$), moderate ($d = .50$), or large ($d = .80$) (J. Cohen, 1969). The meta-analysis mentioned previously included the results of 19 studies (involving 1,683 total participants) of melatonin and concluded that it did indeed significantly improve sleep time although only by about 8 minutes per night on average (Ferracioli-Oda et al., 2013).

Because meta-analyses consider a large number of published, and sometimes unpublished studies, other factors influencing research findings may be evaluated in addition to effect sizes (R. Rosenthal & DiMatteo, 2002). For example, a meta-analysis of altruism found that men were more likely than women to help in risky situations, particularly when others were present (Eagly & Crowley, 1986). Thus, this gender difference might be attributable to the male gender role. Meta-analyses also enable psychologists to compare effect sizes across time, thus assessing the effect of sociocultural change. Two meta-analyses of gender differences in verbal and mathematical ability found that the size of these differences has declined over the years (Hyde et al., 1990; Hyde & Linn, 1988).

Meta-analysis has been applied to many research topics as you will read throughout this textbook, including some that might even relate to your daily life. A meta-analysis of 24 studies involving 3,401 participants found that the infamous commonsense "freshman 15"—a gain of 15 pounds—that supposedly marks the first year of college is inaccurate. The meta-analysis found that there is an average weight gain of about 4 pounds during the first year—closer to a "freshman 5" (Vella-Zarb & Elgar, 2009). And chocolate lovers who are concerned about high blood pressure will appreciate that a meta-analysis of 13 relevant studies found that eating dark chocolate significantly reduces the blood pressure of those in whom it is above normal levels (Ried et al., 2010).

As you read the research studies discussed in later chapters, keep in mind that almost all were analyzed by descriptive statistics, correlational statistics, and inferential statistics. You should also note that statistical significance does not necessarily imply clinical, practical, or social significance (Favreau, 1997; S. J. Lachman, 1993). For instance, participants in an experimental group may differ on the target measure from participants in the control group, but this difference might not be large enough to produce meaningful clinical effects. A small, but statistically significant difference between the experimental and control group might not be large enough to have practical significance outside the laboratory (van Wijk, 2010). Would 8 more minutes of nightly sleep (equivalent to one press of the dreaded snooze button) make a big difference in your life? Some journals that contain reports of research in counseling now require statements of not just statistical significance but also practical or clinical significance (B. Thompson, 2002).

Group Differences Versus Individual Differences

When psychologists report gender, ethnic, or cross-cultural differences in research studies, they are describing group differences on the dependent variable. For example, one study might conclude that boys are more aggressive than girls. This conclusion is based on tests of inferential statistics—the mean for the sample of boys was greater (at a statistically significant level) than the mean for the sample of girls. But a statement about group differences—as in this case of gender differences in aggression—does not mean that the behavior of every male participant differed from that of all the female participants. When frequency distributions of gender, ethnic, or cross-cultural group scores are plotted, there is usually overlap between the two curves. Moreover, most of these studies only included two genders, whereas modern society increasingly views gender on a continuum, with nonbinary or other options typical in current surveys.

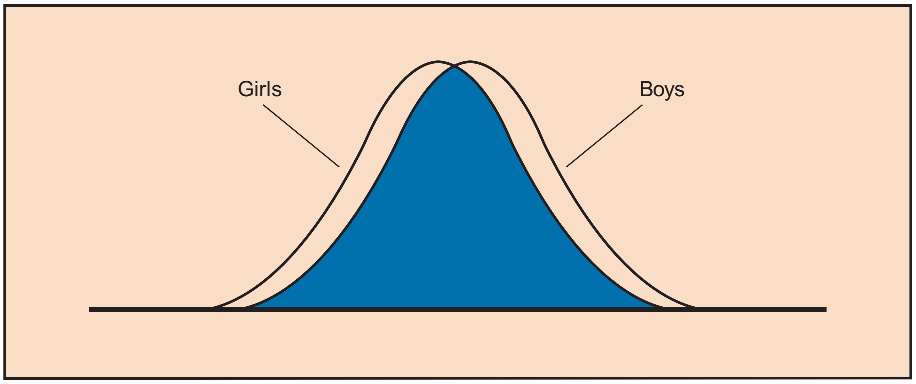


FIGURE 2-3 Statistically Significant Gender Differences

These overlapping curves represent frequency distributions of aggressiveness in a sample of girls and boys. Though these curves represent a statistically significant gender difference in aggressiveness, note that many of the boys and girls did not differ in their aggressiveness (the shaded area of the overlapping curves).

A statistically significant average gender difference in aggression also might still be smaller than individual differences in aggression. As you can see in Figure 2-3, some boys were not very aggressive at all whereas some girls were very aggressive. In fact, the variability of the girls' and boys' scores—the spread of each curve—is greater than the distance between the two group means.

It is important, then, to understand that, although there might be average group differences, it also is likely that there are considerable individual differences. And when individual differences are greater in magnitude than group differences, it is difficult to predict a particular person's behavior on the basis of group differences (we are back to missing the active lava burst in Iceland!). Suppose a researcher reports significant cross-cultural differences between European American and Asian American participants in parenting behaviors. It would be a mistake to conclude from these findings that all European American parents treat their children differently than all Asian Americans. And the difference between any two European American parents is likely to be greater in magnitude than the average cross-cultural difference. This is a reliable adage in cross-cultural research: Differences between individuals within any group are usually larger than average differences between the groups themselves.

Section Review: Statistical Analysis of Research Data

1. What are measures of central tendency?
2. What are measures of variability?
3. What is statistical significance?
4. How does meta-analysis summarize the results of many research studies?

Ethics of Psychological Research

Psychologists must be as concerned with the ethical treatment of their human participants and animal subjects as they are with the quality of their research methods and statistical analyses. Academic psychology departments place a premium on teaching their students the necessity of conducting ethically responsible research (C. B. Fisher et al., 2009).

Ethical Treatment of Research Data

In our hypothetical melatonin experiment, you would want to record your data accurately—even if it contradicted your hypothesis. During the past century, there have been several notorious cases in which researchers in physics, biology, medicine, or psychology have been accused of falsifying their data (R. L. Park, 2008). Chapter 10 discusses a prominent case in psychology in which Sir Cyril Burt, an eminent psychologist, was so intent on demonstrating that intelligence depends on heredity that he apparently misrepresented his research findings (W. H. Tucker, 1997). Though occasional lapses in the ethical treatment of data have provoked controversy, there has been even greater concern about the ethical treatment of people and animals in psychological research.

Ethical Treatment of Human Participants

The first code of ethics for the treatment of human participants in psychological research was developed in 1953, largely in response to the Nuremberg war crimes trials following World War II (C. Miller, 2003). The trials disclosed the cruel medical experiments performed by Nazi physicians on prisoners of war and concentration camp inmates. Today, the U.S. government requires institutions that receive federal research grants to establish a committee—known as an Institutional Review Board (IRB)—that reviews research proposals to ensure the ethical treatment of human participants and animal subjects (McGaha & Korn, 1995).

APA Code of Ethics

The code of ethics of the American Psychological Association (APA), available at https://www.apa.org/ethics/code, was thoroughly revised in 2002 and amended in 2010 and 2016. It contains specific requirements for the treatment of human participants in research and psychological services (testing and therapy). Its key research tenets include

informed consent A bedrock ethical principle in research with humans that safeguards their right to know what the study involves and its potential costs and benefits so they can decide whether or not to participate.

1. The researcher must inform potential participants of all aspects of the research procedure that might influence their decision to participate; this is termed **informed consent**. In the melatonin experiment, you would not be permitted to tell participants that they will receive melatonin and then give them a placebo instead unless they had been informed of that possibility. This requirement, informed consent, can be difficult to ensure if participants are unable to give truly informed consent because they are children (Vitiello, 2008) or have schizophrenia (Beebe & Smith, 2010) or intellectual disabilities such as neurocognitive disorders associated with Alzheimer's disease (Cubit, 2010).
2. Potential participants must not be forced to participate in a research study, which could become a problem with prisoners or hospitalized patients who fear the consequences of refusing to participate (R. Rosenthal, 1995). Though sometimes forcing the individual to undergo therapy, as in the case of adolescents with anorexia nervosa (a disorder marked by self-starvation), can be lifesaving (Manley et al., 2001), it is not ethically permissible in research studies. The laws governing forced medical treatment (not for research purposes) are discussed in Chapter 15.
3. Participants must be permitted to withdraw from a study at any time. Of course, when participants leave, it can adversely affect the study because those who remain might differ from those who drop out (Trice & Ogden, 1987), although there are statistical adjustments to account for this.
4. The researcher must protect the participants from physical harm and mental distress. Certain research practices might raise ethical concerns because of the distress they produce, such as contacting recently bereaved relatives to recruit them to participate in research on mourning (Steeves et al., 2001). If a participant does experience harm or distress, the researcher must try to alleviate it.
5. Information gained from participants must be kept *confidential*. Confidentiality becomes a major issue when participants reveal information that indicates they might be in danger, such as children or adolescents being abused by parents (Wiles et al., 2008).

Deception in Research

Despite this code of ethics, psychologists sometimes confront ethical dilemmas in their treatment of human participants, as in the use of deception to reduce participant bias. Psychologists might purposely misinform participants about the true nature of a study. This deception is of concern, in part, because it violates the ethical norm of informed consent. Recall that the computer dating study by Donn Byrne and colleagues (1970) used deception by falsely claiming that all participants would be matched with partners who shared their attitudes (only half were). Today, for this study to be considered ethical, the researcher would have to demonstrate to the IRB that the experiment could not be conducted without the use of deception and that its potential findings are important enough to justify the use of such deception (C. B. Fisher & Fyrberg, 1994). Moreover, at the completion of the study, each participant would have to be debriefed. In **debriefing** participants, the researcher explains the reasons for the deception and tries to relieve any distress that might have been experienced (Benham, 2008). Some commentators insist that debriefing should be a component of nondeceptive research as well (Sharpe & Faye, 2009).

debriefing A procedure, after the completion of a research study, that informs participants of the true purpose of the study and aims to remove any physical or psychological distress caused by their participation.

Some psychologists worry that deceptive research will make potential participants distrust psychological research (Hertwig & Ortmann, 2008). But arguments against deceptive research have been countered by psychologists who argue that it would be unethical not to conduct deceptive studies that might produce important findings (L. Christensen, 1988). Still others urge psychologists not only to weigh the costs and benefits of using deception but also to inform participants that deception might be used as part of the study (Pittenger, 2002).

Whereas some psychologists argue about the use of deception, others try to settle the debate by using the results of empirical research. In one study, undergraduates who had participated in deceptive experiments rated their experience as more positive than did those who had participated in nondeceptive ones. Moreover, those in deceptive experiments did not rate psychologists as less trustworthy. Any negative emotional effects reported by participants seemed to be relieved by the debriefing process. The researchers concluded that debriefing eliminates any negative effects of deception, perhaps because the participants learn the importance of the research study (S. S. Smith & Richardson, 1983).

Ethical Treatment of Animal Subjects

At the 1986 annual meeting of the American Psychological Association in Washington, D.C., animal rights advocates picketed in the streets and disrupted talks, including one by the prominent psychologist Neal Miller (1909–2002), a defender of the use of animals in psychological research (N. E. Miller, 1985). The conflict between animal rights advocates and psychologists who study animals has long-standing roots. In the early 20th century, animal rights activists attacked the work of leading psychologists, including John B. Watson and G. Stanley Hall. In 1925, in part to blunt these attacks, the American Psychological Association's Committee on Precautions in Animal Experimentation established a code of regulations for the use of animals in research (Dewsbury, 1990).

Scientists Protest in Support of Animal Research

In response to animal rights activists, scientists have begun to demonstrate their support for lifesaving medical research using animals.

Source: Professor David Jentsch. Used with permission.

Animal Rights Versus Animal Welfare

Many *animal rights* advocates oppose all laboratory research using animals, regardless of its scientific merit or practical benefits. Thus, they would oppose testing the effects of melatonin on animal subjects. A survey of demonstrators at an animal rights march in Washington, D.C., in 1990 found that almost 80% of animal rights advocates valued animal life at least as much as human life, and 85% wanted to eliminate all animal research (Plous, 1991). Some animal rights advocates have even vandalized laboratories that conduct animal research to intimidate researchers and interfere with their research (Hadley, 2009).

Conversely, *animal welfare* advocates permit laboratory research on animals as long as the animals are given humane care and the potential benefits of the research outweigh any pain and distress caused to the animals (Wolfensohn & Maguire, 2010). Thus, they would be more likely to approve the use of animals in testing the effects of melatonin on sleep. Bernard Rollin, an ethicist who has tried to resolve the ethical conflict between animal researchers and animal rights advocates, supports animal research but urges that, when in ethical doubt, experimenters should err in favor of the animal (Bekoff et al., 1992).

The current ethical standards of the American Psychological Association and other organizations, such as the Office of Laboratory Animal Welfare, for the treatment of animals are closer to those of animal welfare advocates than to those of animal rights advocates. The standards require that animals be treated with humane care, housed in clean and comfortable enviroments, and given adequate food and water. Researchers also must ensure that their animal subjects experience as little pain and distress as possible. Moreover, all institutions that receive research grants from the United States government must have approval from review boards (similar to IRBs) that judge whether research proposals for experiments using animal subjects meet ethical standards (Holden, 1987a).

Using Animals in Research

But with so many people available, why would psychologists be interested in studying animals at all? Advocates point to a number of benefits derived from animal research, such as the ability to generate findings that can have great benefit to humans with more control of confounds, and perhaps even benefits to the animals themselves (M. E. Carroll & Overmier, 2001).

1. Some psychologists are simply intrigued by animal behavior and wish to learn more about it. To learn about the process of echolocation of prey, you would have to study animals such as bats rather than college students.
2. It is easier to control potential confounding variables that might affect the behavior of an animal. You would be less likely to worry about participant bias effects, for instance, when studying pigeons.
3. Developmental changes across the life span can be studied more efficiently in animals. If you were interested in the effects of the complexity of the early childhood environment on memory in old age, you might take 75 years to complete an experiment using human participants but only 3 years to complete one using rats.
4. Research on animals can generate hypotheses that are then tested using human participants. B. F. Skinner's research on learning in rats and pigeons stimulated research on learning in people.
5. Research on animals can benefit animals themselves.
6. Based on the assumption that animals do not have the same moral rights as people (E. Baldwin, 1993), certain procedures that are not ethically permissible with human participants are ethically permissible under current standards using animal subjects. Thus, if you wanted to conduct an experiment in which you studied the effects of surgically damaging a particular brain structure, you would be limited to the use of animals (Caminiti, 2009). But some critics of animal research note that the use of primates in invasive brain research because they most closely resemble people provides an argument against such research (Crum, 2009). Despite the possible benefits, however, the National Institutes of Health (NIH) decided to phase out much of its research on chimpanzees. As is common with any ethical dilemma, sound arguments can be raised for both sides regarding the use of animals in research. Often there is no resolution that is fully satisfactory to advocates of both sides. We invite you to contemplate and discuss this issue as you deepen your study of psychological science.

We did promise that we would reveal the secret to happiness at the end of this chapter. Your parents, caregivers, or other elders may have told you to get married, have children, and make money. But what does the scientific research reveal? In some cultures, getting married is the best investment you can make in your happiness (G. Kaufman &

Research on Animals May Be Beneficial to Humans and Animals

Source: ProStockStudio/Shutterstock.com.

Taniguchi, 2010); and in a study examining data from 33 different countries, marital satisfaction is independent from (not correlated with) sex, age, duration, education, cultural considerations, or religiosity (Dobrowolska et al., 2020). Unfortunately for many idealistic dreamers in the United States, money does turn out to play a key role too, especially having enough to get out of poverty (Twenge & Cooper, 2020). But in some studies, the effect on happiness of having children is usually close to zero and sometimes even negative (remember a negative correlation coefficient?), meaning that people without children tend to be slightly happier, on average, than those with kids (Kowal et al., 2021). Interestingly, the gap in happiness between parents and nonparents is smaller in countries that provide more resources and social support to families than in countries that provide less assistance, such as the United States (Glass et al., 2016). Makes you appreciate your parents, caregivers, or elders a bit more, perhaps?

Section Review: Ethics of Psychological Research

1. Why has the use of deception in research provoked controversy?
2. What is debriefing in psychological research?
3. How do animal rights and animal welfare advocates differ from one other?

Chapter Summary

Sources of Knowledge

- Psychologists prefer the scientific method to common sense as a source of knowledge.
- The scientific method is based on the assumptions of determinism and skepticism.
- In using the scientific method to conduct research, a psychologist first provides a rationale for the research, then conducts the research study, analyzes the resulting data, and finally, communicates the results to other researchers.
- Replication of research studies is an important component of the scientific research process.

Goals of Scientific Research

- In conducting research, psychologists pursue the goals of description, prediction, control, and explanation.
- Scientific descriptions are systematic and rely on operational definitions.
- Scientific predictions are probabilistic, not certain.
- Scientists exert control over events by manipulating the factors that cause them.
- Scientific explanations state the probable causes of events.

Methods of Psychological Research

- Psychologists use descriptive, correlational, and experimental research methods.
- Descriptive research methods pursue the goal of description through naturalistic observation, case studies, surveys, psychological testing, and archival research.
- Correlational research pursues the goal of prediction by uncovering relationships between variables.
- When using correlational research, psychologists avoid confusing correlation with causation.
- Experimental research pursues the goals of control and explanation by manipulating an independent variable and measuring its effect on a dependent variable.
- Experimenters promote internal validity by controlling confounding variables whose effects might be confused with those of the independent variable.
- Confounding variables might be associated with the experimental situation, the participants in the experiment, or the experimenter.
- Random assignment is used to make the experimental group and control group equivalent before exposing them to the independent variable.
- Experimenters also must control for participant and experimenter bias.
- Another concern of experimenters is external validity—whether their results are generalizable from their participants and settings to other participants and settings.
- Experimenters rely on replication to determine whether their research has external validity.

Statistical Analysis of Research Data

- Psychologists typically make sense of their data by using mathematical techniques called statistics.
- Psychologists use descriptive statistics to summarize data, correlational statistics to determine relationships between variables, and inferential statistics to test their experimental hypotheses.

- Descriptive statistics include measures of central tendency (including the mode, median, and mean) and measures of variability (including the range, variance, and standard deviation).
- Correlational statistics let researchers use the values of one variable to predict the values of another.
- Inferential statistics examine whether numerical differences between experimental and control groups are statistically significant.
- Meta-analysis involves computation of the average effect size across a number of related studies.
- Statistical significance does not necessarily indicate clinical or practical significance. The magnitude of individual differences must be considered when examining group differences.

Ethics of Psychological Research

- Psychologists have ethical codes for the treatment of human participants and animal subjects.
- In research using human participants, researchers must obtain informed consent, not force anyone to participate, let participants withdraw at any time, protect participants from physical harm and mental distress, alleviate any inadvertent harm or distress, and keep information obtained from the participants confidential. The use of deception in research has been an especially controversial issue.
- The use of animals in research also has been controversial. Many animal rights supporters oppose all research on animals. Animal welfare supporters approve of research on animals as long as the animals are treated humanely, and the potential benefits of the research outweigh any pain and distress caused to the animals.
- Most psychologists support the use of animals in research if pain is minimized because of the benefits of such research to both humans and animals.

Key Terms

Sources of Knowledge

determinism (p. 26)
hypothesis (p. 28)
replication (p. 29)
scientific method (p. 28)
skepticism (p. 26)

Goals of Scientific Research

measurement (p. 30)
operational definition (p. 30)
theory (p. 31)

Methods of Psychological Research

archival research (p. 38)
case study (p. 34)
causation (p. 39)
confounding variable (p. 41)
control group (p. 40)
correlation (p. 38)
correlational research (p. 38)
dependent variable (p. 40)
descriptive research (p. 32)
double-blind technique (p. 44)
ethology (p. 33)
experimental group (p. 40)
experimental method (p. 40)
experimenter bias effect (p. 42)
external validity (p. 44)
field experiment (p. 40)
independent variable (p. 40)
internal validity (p. 41)
naturalistic observation (p. 33)
negative correlation (p. 39)
norm (p. 36)
participant bias (p. 42)
placebo (p. 40)
population (p. 35)
positive correlation (p. 39)
psychological test (p. 36)
random assignment (p. 42)
random sampling (p. 35)
reliability (p. 37)
sample (p. 35)
standardization (p. 36)
survey (p. 34)
validity (p. 37)
variable (p. 38)
WEIRD (p. 45)

Statistical Analysis of Research Data

coefficient of correlation (p. 48)
descriptive statistics (p. 46)
inferential statistics (p. 48)
mean (p. 47)
measure of central tendency (p. 46)
measure of variability (p. 48)
median (p. 46)
meta-analysis (p. 50)
mode (p. 46)
range (p. 48)
standard deviation (p. 48)
statistical significance (p. 48)
variance (p. 48)

Ethics of Psychological Research

debriefing (p. 53)
informed consent (p. 52)

Chapter Quiz

Note: Answers for the Chapter Quiz questions are provided at the end of the book.

1. The scientist is governed by an attitude of
a. dualism.
b. cynicism.
c. dogmatism.
d. skepticism.

2. Typically, for a statistical difference between the performances of experimental and control groups to be significant, its probability of occurring by chance must be less than
a. 3%.
b. 5%.
c. 10%.
d. 50%.

3. The prediction that "People high in psychological hardiness will be less likely to become ill than will people low in psychological hardiness" is an example of
a. a fact.
b. a hypothesis.
c. inductive reasoning.
d. an operational definition.

4. Deception is usually used in social psychological research to
a. reduce participant bias.
b. reduce experimenter bias.
c. assess the reactions of people to being fooled.
d. prevent potential participants from finding out about the experimental procedure in advance.

5. The more you smoke, the more likely you are to develop lung cancer. This demonstrates that smoking
a. causes cancer.
b. is caused by cancer.
c. is positively correlated with cancer.
d. is negatively correlated with cancer.

6. If people who score high on a test of mechanical ability perform better on tasks such as fixing a typewriter, building a bookcase, and replacing shock absorbers, this indicates that the test might have
a. reliability.
b. predictive validity.
c. satisfactory norms.
d. no relationship to mechanical ability.

7. You conduct an experiment to investigate the effect of meditation on the level of stress in men and women who vary in their religiosity. The dependent variable in your study is
a. meditation.
b. religiosity.
c. sex of participants.
d. level of stress.

8. The best example of an operational definition would be defining
a. *happy* as "being content with one's life."
b. *beautiful* as "being physically attractive."
c. *expert* as "being knowledgeable in one's own field."
d. *strong* as "being able to bench-press one's body weight."

9. A child psychologist spends 3 hours a week watching and recording the play patterns of 3-year-old children in a nursery school. This is an example of (a)
a. case study.
b. archival research.
c. experimental research.
d. naturalistic observation.

10. The study discussed in the textbook that found that only 5 of 91 persons could identify an intoxicated person demonstrated the shortcomings of
a. science.
b. common sense.
c. deductive reasoning.
d. systematic observation.

11. A psychologist who has designed a personality test administers it to a group of people on two occasions and determines how consistent the performances are. This is a procedure used to assess a test's
a. reliability.
b. validity.
c. norms.
d. criterion.

12. Both participant bias and experimenter bias can be controlled by
a. using the double-blind technique.
b. using more than one independent variable.
c. replicating research studies several times.
d. random assignment of participants to the experimental and control groups.

13. The "Pepsi Challenge" controversy discussed in the textbook revealed that
a. Pepsi tastes better than Coke.
b. Coke tastes better than Pepsi.
c. taste preferences depended on a confounding variable.
d. taste preferences cannot be determined by experiments.

14. Cause is to effect as
a. dependent variable is to independent variable.
b. independent variable is to dependent variable.
c. dependent variable is to confounding variable.
d. confounding variable is to independent variable.

15. The generalizability of research findings is best determined by
a. replication.
b. common sense.
c. archival research.
d. deductive reasoning.

Thought Questions

1. How would a skeptical attitude toward extrasensory perception (ESP), unidentified flying objects (UFOs), and similar topics differ from either a cynical or a gullible attitude?
2. How would the four goals of scientific research influence research on violence?
3. In what way are medical treatments, weather forecasting, horse-race handicapping, college admissions decisions, and psychological child-rearing advice "probabilistic"?
4. Why is the experimental method considered a better means of determining causality than nonexperimental methods?