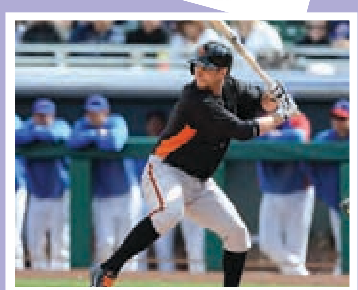
Unit II




Research Methods: Thinking Critically With Psychological Science

Modules

In a difficult moment—after an argument with a loved one, a social embarrassment, or a bad grade—to whom do you turn? For advice and comfort, we often turn to friends and family, or search online. Psychology can also shed insight. Psychologists start with the questions that intrigue all of us: How can we be happier, healthier, and more successful? What can we do to improve our relationships? Why do people act and think as they do? But psychological *science* takes it a step further and uses careful research to separate uninformed opinions from examined conclusions.

Module 4

Top Photo Group/Thinkstock/Getty Images

The Need for Psychological Science

Module Learning Objectives

4-1 Describe how hindsight bias, overconfidence, and the tendency to perceive order in random events illustrate why science-based answers are more valid than those based on intuition and common sense.

4-2 Identify how the three main components of the scientific attitude relate to critical thinking.

4-1 How do hindsight bias, overconfidence, and the tendency to perceive order in random events illustrate why science-based answers are more valid than those based on intuition and common sense?

Some people suppose that psychology merely documents and dresses in jargon what people already know: "So what else is new—you get paid for using fancy methods to prove what everyone knows?" Others place their faith in human intuition: "Buried deep within each and every one of us, there is an instinctive, heart-felt awareness that provides—if we allow it to—the most reliable guide," offered Prince Charles (2000).

Prince Charles has much company, judging from the long list of pop psychology books on "intuitive managing," "intuitive trading," and "intuitive healing." Today's psychological science does document a vast intuitive mind. As we will see, our thinking, memory, and attitudes operate on two levels—conscious and unconscious—with the larger part operating automatically, off-screen. Like jumbo jets, we fly mostly on autopilot.

The limits of intuition Personnel interviewers tend to be overconfident of their gut feelings about job applicants. Their confidence stems partly from their recalling cases where their favorable impression proved right, and partly from their ignorance about rejected applicants who succeeded elsewhere.

Marilyn Angel Wynn/Getty Images

So, are we smart to listen to the whispers of our inner wisdom, to simply trust "the force within"? Or should we more often be subjecting our intuitive hunches to skeptical scrutiny?

This much seems certain: We often underestimate intuition's perils. My geographical intuition tells me that Reno is east of Los Angeles, that Rome is south of New York, that Atlanta is east of Detroit. But I am wrong, wrong, and wrong.

Modules to come will show that experiments have found people greatly overestimating their lie detection accuracy, their eyewitness recollections, their interviewee assessments, their risk predictions, and their stock-picking talents. As a Nobel Prize-winning physicist explained, "The first principle is that you must not fool yourself—and you are the easiest person to fool" (Feynman, 1997).

Indeed, observed novelist Madeleine L'Engle, "The naked intellect is an extraordinarily inaccurate instrument" (1973). Three phenomena—*hindsight bias, judgmental overconfidence, and our tendency to perceive patterns in random events*—illustrate why we cannot rely solely on intuition and common sense.

"Those who trust in their own wits are fools." -Proverbs 28:26

Did We Know It All Along? Hindsight Bias

Consider how easy it is to draw the bull's eye *after* the arrow strikes. After the stock market drops, people say it was "due for a correction." After the football game, we credit the coach if a "gutsy play" wins the game, and fault the coach for the "stupid play" if it doesn't. After a war or an election, its outcome usually seems obvious. Although history may therefore seem like a series of inevitable events, the actual future is seldom foreseen. No one's diary recorded, "Today the Hundred Years War began."

"Life is lived forwards, but understood backwards." -Philosopher Søren Kierkegaard, 1813–1855

This **hindsight bias** (also known as the *I-knew-it-all-along phenomenon*) is easy to demonstrate: Give half the members of a group some purported psychological finding, and give the other half an opposite result. Tell the first group, "Psychologists have found that separation weakens romantic attraction. As the saying goes, 'Out of sight, out of mind.'" Ask them to imagine why this might be true. Most people can, and nearly all will then view this true finding as unsurprising.

hindsight bias the tendency to believe, after learning an outcome, that one would have foreseen it. (Also known as the *I-knew-it-all-along phenomenon.*)

Tell the second group the opposite, "Psychologists have found that separation strengthens romantic attraction. As the saying goes, 'Absence makes the heart grow fonder.'" People given this untrue result can also easily imagine it, and most will also see it as unsurprising. When two opposite findings both seem like common sense, there is a problem.

Such errors in our recollections and explanations show why we need psychological research. Just asking people how and why they felt or acted as they did can sometimes be misleading—not because common sense is usually wrong, but because common sense more easily describes what *has* happened than what *will* happen. As physicist Niels Bohr reportedly said, "Prediction is very difficult, especially about the future."

"Anything seems commonplace, once explained." -Dr. Watson to Sherlock Holmes

Some 100 studies have observed hindsight bias in various countries and among both children and adults (Blank et al., 2007). Nevertheless, our intuition is often right. As Yogi Berra once said, "You can observe a lot by watching." (We have Berra to thank for other gems, such as "Nobody ever comes here—it's too crowded," and "If the people don't want to come out to the ballpark, nobody's gonna stop 'em.") Because we're all behavior watchers, it would be

REUTERS/U.S. Coast Guard/Handout

Hindsight bias When drilling the Deepwater Horizon oil well in 2010, BP employees took some shortcuts and ignored some warning signs, without intending to harm the environment or their company's reputation. *After* the resulting Gulf oil spill, with the benefit of 20/20 hindsight, the foolishness of those judgments became obvious.

AP® Exam Tip

It is quite common for multiple-choice questions on the AP® exam to test your knowledge of "media myths." Pay particular attention when psychological findings run counter to "common sense."

surprising if many of psychology's findings had *not* been foreseen. Many people believe that love breeds happiness, and they are right (we have what Module 40 calls a deep "need to belong"). Indeed, note Daniel Gilbert, Brett Pelham, and Douglas Krull (2003), "good ideas in psychology usually have an oddly familiar quality, and the moment we encounter them we feel certain that we once came close to thinking the same thing ourselves and simply failed to write it down." Good ideas are like good inventions; once created, they seem obvious. (Why did it take so long for someone to invent suitcases on wheels and Post-it Notes?)

But sometimes our intuition, informed by countless casual observations, has it wrong. In later modules we will see how research has overturned popular ideas—that familiarity breeds contempt, that dreams predict the future, and that most of us use only 10 percent of our brain. (See also **TABLE 4.1**.) We will also see how it has surprised us with discoveries about how the brain's chemical messengers control our moods and memories, about other animals' abilities, and about the effects of stress on our capacity to fight disease.

Table 4.1 True or False?
Psychological research discussed in modules to come will either confirm or refute each of these statements (adapted, in part, from Furnham et al., 2003). Can you predict which of these popular ideas have been confirmed and which refuted? (Check your answers at the bottom of this table.)
1. If you want to teach a habit that persists, reward the desired behavior every time, not just intermittently (see Module 27).
2. Patients whose brains are surgically split down the middle survive and function much as they did before the surgery (see Module 13).
3. Traumatic experiences, such as sexual abuse or surviving the Holocaust, are typically "repressed" from memory (see Module 33).
4. Most abused children do *not* become abusive adults (see Module 50).
5. Most infants recognize their own reflection in a mirror by the end of their first year (see Module 47).
6. Adopted siblings usually do not develop similar personalities, even though they are reared by the same parents (see Module 14).
7. Fears of harmless objects, such as flowers, are just as easy to acquire as fears of potentially dangerous objects, such as snakes (see Module 15).
8. Lie detection tests often lie (see Module 41).
9. The brain remains active during sleep (see Modules 22–23).
Answers: 1. F, 2. T, 3. F, 4. T, 5. F, 6. T, 7. F, 8. T, 9. T

Overconfidence

We humans tend to think we know more than we do. Asked how sure we are of our answers to factual questions *(Is Boston north or south of Paris?)*, we tend to be more confident than correct.[1] Or consider these three anagrams, which Richard Goranson (1978) asked people to unscramble:

WREAT → WATER

ETRYN → ENTRY

GRABE → BARGE

About how many seconds do you think it would have taken you to unscramble each of these? Did hindsight influence you? Knowing the answers tends to make us overconfident—surely the solution would take only 10 seconds or so, when in reality the average problem solver spends 3 minutes, as you also might, given a similar anagram without the solution: OCHSA.[2]

Are we any better at predicting social behavior? University of Pennsylvania psychologist Philip Tetlock (1998, 2005) collected more than 27,000 expert predictions of world events, such as the future of South Africa or whether Quebec would separate from Canada. His repeated finding: These predictions, which experts made with 80 percent confidence on average, were right less than 40 percent of the time. Nevertheless, even those who erred maintained their confidence by noting they were "almost right." "The Québécois separatists *almost* won the secessionist referendum."

Overconfidence in history:

"We don't like their sound. Groups of guitars are on their way out." -Decca Records, in turning down a recording contract with the Beatles in 1962

"Computers in the future may weigh no more than 1.5 tons." -*Popular Mechanics*, 1949

"They couldn't hit an elephant at this distance." -General John Sedgwick just before being killed during a U.S. Civil War battle, 1864

"The telephone may be appropriate for our American cousins, but not here, because we have an adequate supply of messenger boys." -British expert group evaluating the invention of the telephone

Perceiving Order in Random Events

In our natural eagerness to make sense of our world—what poet Wallace Stevens called our "rage for order"—we are prone to perceive patterns. People see a face on the moon, hear Satanic messages in music, perceive the Virgin Mary's image on a grilled cheese sandwich. Even in random data we often find order, because—here's a curious fact of life—*random sequences often don't look random* (Falk et al., 2009; Nickerson, 2002, 2005). Consider a random coin flip: If someone flipped a coin six times, which of the following sequences of heads (H) and tails (T) would be most likely: HHHTTT or HTTHTH or HHHHHH?

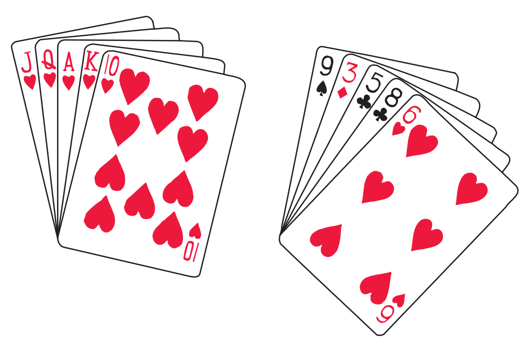
Daniel Kahneman and Amos Tversky (1972) found that most people believe HTTHTH would be the most likely random sequence. Actually, all three are equally likely (or, you might say, equally unlikely). A poker hand of 10 through ace, all of hearts, would seem extraordinary; actually, it would be no more or less likely than any other specific hand of cards (**FIGURE 4.1**).

In actual random sequences, patterns and streaks (such as repeating digits) occur more often than people expect (Oskarsson et al., 2009). To demonstrate this phenomenon for myself, I flipped a coin 51 times, with these results:

1. H	10. T	19. H	28. T	37. T	46. H
2. T	11. T	20. H	29. H	38. T	47. H
3. T	12. H	21. T	30. T	39. H	48. T
4. T	13. H	22. T	31. T	40. T	49. T
5. H	14. T	23. H	32. T	41. H	50. T
6. H	15. T	24. T	33. T	42. H	51. T
7. H	16. H	25. T	34. T	43. H	
8. T	17. T	26. T	35. T	44. H	
9. T	18. T	27. H	36. H	45. T	

Looking over the sequence, patterns jump out: Tosses 10 to 22 provided an almost perfect pattern of pairs of tails followed by pairs of heads. On tosses 30 to 38 I had a "cold hand," with only one head in nine tosses. But my fortunes immediately reversed with a "hot hand"—seven heads out of the next nine tosses. Similar streaks happen, about as often as one would expect in random sequences, in basketball shooting, baseball hitting, and mutual fund stock pickers' selections (Gilovich et al., 1985; Malkiel, 2007; Myers, 2002). These sequences often don't look random and so are overinterpreted. ("When you're hot, you're hot!")

Figure 4.1
Two random sequences Your chances of being dealt either of these hands are precisely the same: 1 in 2,598,960.



[1] Boston is south of Paris.
[2] The anagram solution: CHAOS.

DORA: Okay that FTN 1 is now on facing page from where it occurs in text?

What explains these streaky patterns? Was I exercising some sort of paranormal control over my coin? Did I snap out of my tails funk and get in a heads groove? No such explanations are needed, for these are the sorts of streaks found in any random data. Comparing each toss to the next, 23 of the 50 comparisons yielded a changed result—just the sort of near 50-50 result we expect from coin tossing. Despite seeming patterns, the outcome of one toss gives no clue to the outcome of the next.

However, some happenings seem so extraordinary that we struggle to conceive an ordinary, chance-related explanation (as applies to our coin tosses). In such cases, statisticians often are less mystified. When Evelyn Marie Adams won the New Jersey lottery *twice,* newspapers reported the odds of her feat as 1 in 17 trillion. Bizarre? Actually, 1 in 17 trillion are indeed the odds that a given person who buys a single ticket for two New Jersey lotteries will win both times. And given the millions of people who buy U.S. state lottery tickets, statisticians Stephen Samuels and George McCabe (1989) reported, it was "practically a sure thing" that someday, somewhere, someone would hit a state jackpot twice. Indeed, said fellow statisticians Persi Diaconis and Frederick Mosteller (1989), "with a large enough sample, any outrageous thing is likely to happen." An event that happens to but 1 in 1 billion people every day occurs about 7 times a day, 2500 times a year.

"The really unusual day would be one where nothing unusual happens." -Statistician Persi Diaconis (2002)

The point to remember: Hindsight bias, overconfidence, and our tendency to perceive patterns in random events often lead us to overestimate our intuition. But scientific inquiry can help us sift reality from illusion.

The Scientific Attitude: Curious, Skeptical, and Humble

4-2 How do the scientific attitude's three main components relate to critical thinking?

Underlying all science is, first, a hard-headed *curiosity,* a passion to explore and understand without misleading or being misled. Some questions *(Is there life after death?)* are beyond science. Answering them in any way requires a leap of faith. With many other ideas *(Can some people demonstrate ESP?),* the proof is in the pudding. Let the facts speak for themselves.

Magician James Randi has used this *empirical approach* when testing those claiming to see auras around people's bodies:

Randi: Do you see an aura around my head?

Aura-seer: Yes, indeed.

Randi: Can you still see the aura if I put this magazine in front of my face?

Aura-seer: Of course.

Randi: Then if I were to step behind a wall barely taller than I am, you could determine my location from the aura visible above my head, right?

Randi told me that no aura-seer has agreed to take this simple test.

No matter how sensible-seeming or wild an idea, the smart thinker asks: *Does it work?* When put to the test, can its predictions be confirmed? Subjected to such scrutiny, crazy-sounding ideas sometimes find support. During the 1700s, scientists scoffed at the notion that meteorites had extraterrestrial origins. When two Yale scientists challenged the conventional opinion, Thomas Jefferson jeered, "Gentlemen, I would rather believe that those two Yankee professors would lie than to believe that stones fell from Heaven." Sometimes scientific inquiry turns jeers into cheers.

More often, science becomes society's garbage disposal, sending crazy-sounding ideas to the waste heap, atop previous claims of perpetual motion machines, miracle cancer cures, and out-of-body travels into centuries past. To sift reality from fantasy, sense from nonsense, therefore requires a scientific attitude: being skeptical but not cynical, open but not gullible.

The Amazing Randi The magician James Randi exemplifies skepticism. He has tested and debunked supposed psychic phenomena.

AP Photo/Alan Diaz

"To believe with certainty," says a Polish proverb, "we must begin by doubting." As scientists, psychologists approach the world of behavior with a *curious skepticism,* persistently asking two questions: *What do you mean? How do you know?*

When ideas compete, skeptical testing can reveal which ones best match the facts. Do parental behaviors determine children's sexual orientation? Can astrologers predict your future based on the position of the planets at your birth? Is electroconvulsive therapy (delivering an electric shock to the brain) an effective treatment for severe depression? As we will see, putting such claims to the test has led psychological scientists to answer *No* to the first two questions and *Yes* to the third.

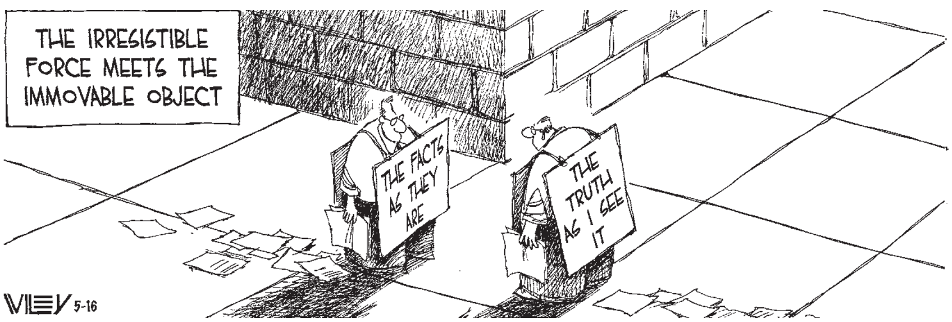
"I'm a skeptic not because I do not want to believe but because I want to *know*. I believe that the truth is out there. But how can we tell the difference between what we would like to be true and what is actually true? The answer is science." -Michael Shermer, "I Want to Believe," *Scientific American*, 2009

Putting a scientific attitude into practice requires not only curiosity and skepticism but also *humility*—an awareness of our own vulnerability to error and an openness to surprises and new perspectives. In the last analysis, what matters is not my opinion or yours, but the truths nature reveals in response to our questioning. If people or other animals don't behave as our ideas predict, then so much the worse for our ideas. This humble attitude was expressed in one of psychology's early mottos: "The rat is always right."

Historians of science tell us that these three attitudes—curiosity, skepticism, and humility—helped make modern science possible. Some deeply religious people today may view science, including psychological science, as a threat. Yet, many of the leaders of the scientific revolution, including Copernicus and Newton, were deeply religious people acting on the idea that "in order to love and honor God, it is necessary to fully appreciate the wonders of his handiwork" (Stark, 2003a,b).

"My deeply held belief is that if a god anything like the traditional sort exists, our curiosity and intelligence are provided by such a god. We would be unappreciative of those gifts . . . if we suppressed our passion to explore the universe and ourselves." -Carl Sagan, *Broca's Brain*, 1979

Non Sequitur



Of course, scientists, like anyone else, can have big egos and may cling to their preconceptions. Nevertheless, the ideal of curious, skeptical, humble scrutiny of competing ideas unifies psychologists as a community as they check and recheck one another's findings and conclusions.

Critical Thinking

The scientific attitude prepares us to think smarter. Smart thinking, called **critical thinking,** examines assumptions, assesses the source, discerns hidden values, confirms evidence, and assesses conclusions. Whether reading a news report or listening to a conversation, critical thinkers ask questions. Like scientists, they wonder: How do they know that? What is this person's agenda? Is the conclusion based on anecdote and gut feelings, or on evidence? Does the evidence justify a cause-effect conclusion? What alternative explanations are possible?

critical thinking thinking that does not blindly accept arguments and conclusions. Rather, it examines assumptions, assesses the source, discerns hidden values, evaluates evidence, and assesses conclusions.

Critical thinking, informed by science, helps clear the colored lenses of our biases. Consider: Does climate change threaten our future, and, if so, is it human-caused? In 2009, climate-action advocates interpreted an Australian heat wave and dust storms as evidence of climate change. In 2010, climate-change skeptics perceived North American bitter cold and East Coast blizzards as discounting global warming. Rather than having their understanding

> "The real purpose of the scientific method is to make sure Nature hasn't misled you into thinking you know something you don't actually know." -Robert M. Pirsig, *Zen and the Art of Motorcycle Maintenance*, 1974

of climate change swayed by today's weather, or by their own political views, critical thinkers say, "Show me the evidence." Over time, is the Earth actually warming? Are the polar ice caps melting? Are vegetation patterns changing? And is human activity spewing gases that would lead us to expect such changes? When contemplating such issues, critical thinkers will consider the credibility of sources. They will look at the evidence (*"Do the facts support them, or are they just makin' stuff up?")*. They will recognize multiple perspectives. And they will expose themselves to news sources that challenge their preconceived ideas.

Has psychology's critical inquiry been open to surprising findings? The answer, as ensuing modules illustrate, is plainly *Yes*. Believe it or not, massive losses of brain tissue early in life may have minimal long-term effects (see Module 12). Within days, newborns can recognize their mother's odor and voice (see Module 45). After brain damage, a person may be able to learn new skills yet be unaware of such learning (see Modules 31–33). Diverse groups—men and women, old and young, rich and middle class, those with disabilities and without—report roughly comparable levels of personal happiness (see Module 83).

And has critical inquiry convincingly debunked popular presumptions? The answer, as ensuing modules also illustrate, is again *Yes.* The evidence indicates that sleepwalkers are *not* acting out their dreams (see Module 24). Our past experiences are *not* all recorded verbatim in our brains; with brain stimulation or hypnosis, one *cannot* simply "hit the replay button" and relive long-buried or repressed memories (see Module 33). Most people do *not* suffer from unrealistically low self-esteem, and high self-esteem is not all good (see Module 59). Opposites do *not* generally attract (see Module 79). In each of these instances and more, what has been learned is not what is widely believed.

Before You Move On

▶ ASK YOURSELF

How might critical thinking help us assess someone's interpretations of people's dreams or their claims to communicate with the dead?

▶ TEST YOURSELF

How does the scientific attitude contribute to critical thinking?

Answers to the Test Yourself questions can be found in Appendix E at the end of the book.

Module 4 Review

4-1 How do hindsight bias, overconfidence, and the tendency to perceive order in random events illustrate why science-based answers are more valid than those based on intuition and common sense?

- *Hindsight bias* (also called the "I-knew-it-all-along phenomenon") is the tendency to believe, after learning an outcome, that we would have foreseen it.
- Overconfidence in our judgments results partly from our bias to seek information that confirms them.
- These tendencies, plus our eagerness to perceive patterns in random events, lead us to overestimate our intuition.

Although limited by the testable questions it can address, scientific inquiry can help us overcome our intuition's biases and shortcomings.

4-2 How do the scientific attitude's three main components relate to critical thinking?

- The scientific attitude equips us to be curious, skeptical, and humble in scrutinizing competing ideas or our own observations.
- This attitude carries into everyday life as *critical thinking,* which puts ideas to the test by examining assumptions, assessing the source, discerning hidden values, evaluating evidence, and assessing conclusions.

Multiple-Choice Questions

1. After the student council election, a friend tells you he has known for weeks who would be elected president. What does this seem to illustrate?

a. Skepticism
b. Critical thinking
c. Hindsight bias
d. Overconfidence
e. Perceiving order in random events

2. While taking a standardized test with randomly scrambled answers, you notice that your last four answers have been "c." Which of the following is true concerning the probability of the next answer being "c"?

a. It is higher. Once a streak begins, it is likely to last for a while.
b. It is lower. Since answers are distributed randomly, "c" answers become less common.
c. It is unaffected by previous answers. It is as likely to be "c" as any other answer.
d. You should check your previous answers. Four "c's" in a row is impossible.
e. It is higher. Test constructors trick students by keeping the same answer many times in a row.

3. What do we call the tendency to exaggerate the correctness or accuracy of our beliefs and predictions prior to testing?

a. Hindsight bias
b. Overconfidence
c. Critical thinking
d. Skepticism
e. Reliability

4. Which of the following is an example of hindsight bias?

a. Tom is certain that electric cars will represent 80 percent of vehicles in twenty years and only reads research studies that support his hypothesis.
b. Liza underestimates how much time it will take her to finish writing her college application essays and as a result fails to meet an important deadline.
c. Experts predicting world events with 80 percent confidence turned out to be correct less than 40 percent of the time.
d. Marcy cannot recognize a definition on a flashcard. After turning the card over and viewing the term, she tells herself she knew what the answer was all along.
e. Dr. Grace overestimates how effectively her new treatment method works because she fails to seek out any evidence refuting her theory.

Practice FRQs

1. Name the three components of the scientific attitude. Provide an example to show how each component contributes to the investigation of competing ideas in psychology.

Answer

1 point: Curiosity, or passion to explore, leads us to questions we want to investigate. Any examples of such questions will serve (For example, Does more money make us happier? Is schizophrenia inherited?).

1 point: Skepticism keeps us from accepting ideas without sound support. The work of The Amazing Randi would be a good example here.

1 point: Humility keeps us open to the possibility of changing our ideas when they are not supported by the data. For example, "the rat is always right."

2. Aziz has read that handwriting reveals important details about personality. Explain how each component of the scientific attitude can help Aziz investigate the accuracy of the information he has read about handwriting analysis.

(3 points)

Module 5

The Scientific Method and Description

Module Learning Objectives

5-1 Describe how theories advance psychological science.

5-2 Describe how psychologists use case studies, naturalistic observation, and surveys to observe and describe behavior, and explain the importance of random sampling.

AP® Exam Tip

As you read this module, keep in mind that the scientific method is a set of principles and procedures, not a list of facts. You will be expected to understand how the science of psychology is done, not just what it has discovered.

Psychologists arm their scientific attitude with the *scientific method*—a self-correcting process for evaluating ideas with observation and analysis. In its attempt to describe and explain human nature, psychological science welcomes hunches and plausible-sounding theories. And it puts them to the test. If a theory works—if the data support its predictions—so much the better for that theory. If the predictions fail, the theory will be revised or rejected.

The Scientific Method

5-1 How do theories advance psychological science?

theory an explanation using an integrated set of principles that organizes observations and predicts behaviors or events.

Chatting with friends and family, we often use theory to mean "mere hunch." In science, a **theory** *explains* behaviors or events by offering ideas that *organize* what we have observed. By organizing isolated facts, a theory simplifies. By linking facts with deeper principles, a theory offers a useful summary. As we connect the observed dots, a coherent picture emerges.

A theory about the effects of sleep on memory, for example, helps us organize countless sleep-related observations into a short list of principles. Imagine that we observe over and over that people with good sleep habits tend to answer questions correctly in class, and they do well at test time. We might therefore theorize that sleep improves memory. So far so good: Our principle neatly summarizes a list of facts about the effects of a good night's sleep on memory.

hypothesis a testable prediction, often implied by a theory.

Yet no matter how reasonable a theory may sound—and it does seem reasonable to suggest that sleep could improve memory—we must put it to the test. A good theory produces testable predictions, called **hypotheses.** Such predictions specify what results (what behaviors or events) would support the theory and what results would cast doubt on the theory. To test our theory about the effects of sleep on memory, our hypothesis might be that when sleep deprived, people will remember less from the day before. To test that hypothesis, we might assess how well people remember course materials they studied before a good night's sleep, or before a shortened night's sleep (**FIGURE 5.1**). The results will either confirm our theory or lead us to revise or reject it.

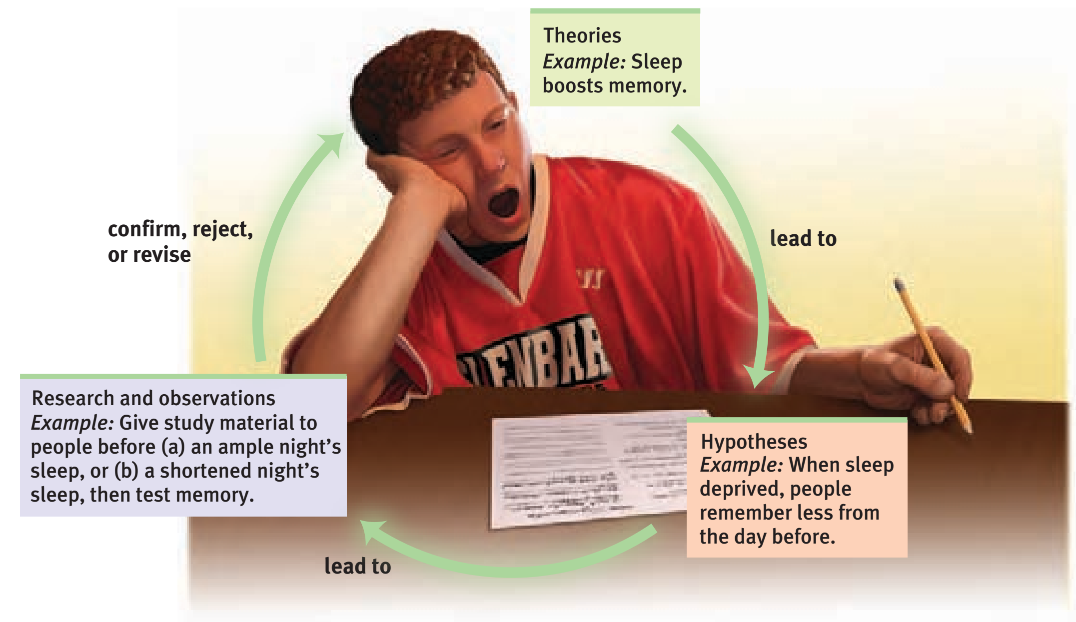


Figure 5.1
The scientific method A self-correcting process for asking questions and observing nature's answers.

Our theories can bias our observations. Having theorized that better memory springs from more sleep, we may see what we expect: We may perceive sleepy people's comments as less insightful. Perhaps you are aware of students who, because they have developed an excellent reputation, can now do no wrong in the eyes of teachers. If they're in the hall during class, nobody worries. Other students can do no good. Because they have behaved badly in the past, even their positive behaviors are viewed suspiciously.

As a check on their biases, psychologists use **operational definitions** when they report their studies. "Sleep deprived," for example, may be defined as "X hours less" than the person's natural sleep. Unlike dictionary definitions, operational definitions describe concepts with precise procedures or measures. These exact descriptions will allow anyone to **replicate** (repeat) the research. Other people can then re-create the study with different participants and in different situations. If they get similar results, we can be confident that the findings are reliable.

operational definition a carefully worded statement of the exact procedures (operations) used in a research study. For example, *human intelligence* may be operationally defined as what an intelligence test measures.

replication repeating the essence of a research study, usually with different participants in different situations, to see whether the basic finding extends to other participants and circumstances.

Let's summarize. A good theory:

- effectively *organizes* a range of self-reports and observations.
- leads to clear *hypotheses* (predictions) that anyone can use to check the theory.
- often stimulates research that leads to a revised theory which better organizes and predicts what we know. Or, our research may be replicated and supported by similar findings. (This has been the case for sleep and memory studies, as you will see in Module 24.)

We can test our hypotheses and refine our theories in several ways.

- *Descriptive* methods describe behaviors, often by using case studies, surveys, or naturalistic observations.
- *Correlational* methods associate different factors, or *variables.* (You'll see the word *variable* often in descriptions of research. It refers to anything that contributes to a result.)
- *Experimental* methods manipulate variables to discover their effects.

To think critically about popular psychology claims, we need to understand the strengths and weaknesses of these methods.

Description

5-2 How do psychologists use case studies, naturalistic observation, and surveys to observe and describe behavior, and why is random sampling important?

The starting point of any science is description. In everyday life, we all observe and describe people, often drawing conclusions about why they act as they do. Professional psychologists do much the same, though more objectively and systematically, through

- *case studies* (analyses of special individuals).
- *naturalistic observation* (watching and recording the natural behavior of many individuals).
- *surveys* and interviews (by asking people questions).

case study a descriptive technique in which one individual or group is studied in depth in the hope of revealing universal principles.

naturalistic observation observing and recording behavior in naturally occurring situations without trying to manipulate and control the situation.

The Case Study

Psychologists use the **case study,** which is among the oldest research methods, to examine one individual or group in depth in the hope of revealing things true of all of us. Some examples: Much of our early knowledge about the brain came from case studies of individuals who suffered a particular impairment after damage to a certain brain region. Jean Piaget taught us about children's thinking through case studies in which he carefully observed and questioned individual children. Studies of only a few chimpanzees have revealed their capacity for understanding and language. Intensive case studies are sometimes very revealing. They show us what *can* happen, and they often suggest directions for further study.

"'Well my dear,' said Miss Marple, 'human nature is very much the same everywhere, and of course, one has opportunities of observing it at closer quarters in a village.'" -Agatha Christie, *The Tuesday Club Murders*, 1933

But individual cases may mislead us if the individual is atypical. Unrepresentative information can lead to mistaken judgments and false conclusions. Indeed, anytime a researcher mentions a finding *("Smokers die younger: ninety-five percent of men over 85 are nonsmokers")* someone is sure to offer a contradictory anecdote *("Well, I have an uncle who smoked two packs a day and lived to 89")*. Dramatic stories and personal experiences (even psychological case examples) command our attention and are easily remembered. Journalists understand that, and so begin an article about bank foreclosures with the sad story of one family put out of their house, not with foreclosure statistics. Stories move us. But stories can mislead. Which of the following do you find more memorable? (1) "In one study of 1300 dream reports concerning a kidnapped child, only 5 percent correctly envisioned the child as dead" (Murray & Wheeler, 1937). (2) "I know a man who dreamed his sister was in a car accident, and two days later she died in a head-on collision!" Numbers can be numbing, but the plural of *anecdote* is not *evidence.* As psychologist Gordon Allport (1954, p. 9) said, "Given a thimbleful of [dramatic] facts we rush to make generalizations as large as a tub."

The point to remember: Individual cases can suggest fruitful ideas. What's true of all of us can be glimpsed in any one of us. But to discern the general truths that cover individual cases, we must answer questions with other research methods.

Juniors Bildarchiv/F355/Alamy

Freud and Little Hans Sigmund Freud's case study of 5-year-old Hans' extreme fear of horses led Freud to his theory of childhood sexuality. He conjectured that Hans felt unconscious desire for his mother, feared castration by his rival father, and then transferred this fear into his phobia about being bitten by a horse. As Module 56 will explain, today's psychological science discounts Freud's theory of childhood sexuality but acknowledges that much of the human mind operates outside our conscious awareness.

Naturalistic Observation

A second descriptive method records behavior in natural environments. These **naturalistic observations** range from watching chimpanzee societies in the jungle, to unobtrusively videotaping (and later systematically analyzing) parent-child interactions in different cultures, to recording racial differences in students' self-seating patterns in a school cafeteria.

Like the case study, naturalistic observation does not *explain* behavior. It *describes* it. Nevertheless, descriptions can be revealing. We once thought, for example, that only humans use tools. Then naturalistic observation revealed that chimpanzees sometimes insert a stick in a termite mound and withdraw it, eating the stick's load of termites.

Such unobtrusive naturalistic observations paved the way for later studies of animal thinking, language, and emotion, which further expanded our understanding of our fellow animals. "Observations, made in the natural habitat, helped to show that the societies and behavior of animals are far more complex than previously supposed," chimpanzee observer Jane Goodall noted (1998). Thanks to researchers' observations, we know that chimpanzees and baboons use deception. Psychologists Andrew Whiten and Richard Byrne (1988) repeatedly saw one young baboon pretending to have been attacked by another as a tactic to get its mother to drive the other baboon away from its food. The more developed a primate species' brain, the more likely it is that the animals will display deceptive behaviors (Byrne & Corp, 2004).

Martin Harvey/Getty Images

A natural observer Chimpanzee researcher Frans de Waal (2005) reported, "I am a born observer. . . . When picking a seat in a restaurant I want to face as many tables as possible. I enjoy following the social dynamics—love, tension, boredom, antipathy—around me based on body language, which I consider more informative than the spoken word. Since keeping track of others is something I do automatically, becoming a fly on the wall of an ape colony came naturally to me."

Naturalistic observations also illuminate human behavior. Here are four findings you might enjoy.

- *A funny finding.* We humans laugh 30 times more often in social situations than in solitary situations. (Have you noticed how seldom you laugh when alone?) As we laugh, 17 muscles contort our mouth and squeeze our eyes, and we emit a series of 75-millisecond vowel-like sounds, spaced about one-fifth of a second apart (Provine, 2001).
- *Sounding out students.* What, really, are college psychology students saying and doing during their everyday lives? To find out, researchers equipped 52 such students from the University of Texas with electronic recorders (Mehl & Pennebaker, 2003). For up to four days, the recorders captured 30 seconds of the students' waking hours every 12.5 minutes, thus enabling the researchers to eavesdrop on more than 10,000 half-minute life slices by the end of the study. On what percentage of the slices do you suppose they found the students talking with someone? What percentage captured the students at a computer? The answers: 28 and 9 percent. (What percentage of *your* waking hours are spent in these activities?)
- *What's on your mind?* To find out what was on the mind of their University of Nevada, Las Vegas, students, researchers gave them beepers (Heavey & Hurlburt, 2008). On a half-dozen occasions, a beep interrupted students' daily activities, signaling them to pull out a notebook and record their inner experience at that moment. When the researchers later coded the reports in categories, they found five common forms of inner experience (**TABLE 5.1** on the next page).
- *Culture, climate, and the pace of life.* Naturalistic observation also enabled researchers to compare the pace of life in 31 countries (Levine & Norenzayan, 1999). (Their operational definition of *pace of life* included walking speed, the speed with which postal clerks completed a simple request, and the accuracy of public clocks.) Their conclusion: Life is fastest paced in Japan and Western Europe, and slower paced in economically less-developed countries. People in colder climates also tend to live at a faster pace (and are more prone to die from heart disease).

Courtesy of Matthias Mehl

Naturalistic observation Researchers at the University of Texas used electronic recorders to sample naturally occurring slices of daily life.

Table 5.1 A Penny for Their Thoughts: The Inner Experience of University Students*

Inner Experience	Example	Frequency
Inner speech	Susan was saying to herself, "I've got to get to class."	26%
Inner seeing	Paul was imagining the face of a best friend, including her neck and head.	34%
Unsymbolized thinking	Alphonse was wondering whether the workers would drop the bricks.	22%
Feeling	Courtney was experiencing anger and its physical symptoms.	26%
Sensory awareness	Fiona was feeling the cold breeze on her cheek and her hair moving.	22%

* More than one experience could occur at once.

Naturalistic observation offers interesting snapshots of everyday life, but it does so without controlling for all the variables that may influence behavior. It's one thing to observe the pace of life in various places, but another to understand what makes some people walk faster than others.

The Survey

survey a technique for ascertaining the self-reported attitudes or behaviors of a particular group, usually by questioning a representative, random sample of the group.

A **survey** looks at many cases in less depth. Researchers do surveys when wanting to estimate, from a representative sample of people, the attitudes or reported behaviors of a whole population. Questions about everything from cell-phone use to political opinions are put to the public. In recent surveys,

- half of all Americans reported experiencing more happiness and enjoyment than worry and stress on the previous day (Gallup, 2010).
- online Canadians reported using new forms of electronic communication and thus receiving 35 percent fewer e-mails in 2010 than 2008 (Ipsos, 2010a).
- 1 in 5 people across 22 countries reported believing that alien beings have come to Earth and now walk among us disguised as humans (Ipsos, 2010b).
- 68 percent of all humans—some 4.6 billion people—say that religion is important in their daily lives (Diener et al., 2011).

But asking questions is tricky, and the answers often depend on the ways questions are worded and respondents are chosen.

WORDING EFFECTS

As we will see in Module 35, even subtle changes in the order or wording of questions—the way we *frame* a question—can have major effects. People are much more approving of "aid to the needy" than of "welfare," of "affirmative action" than of "preferential treatment," of "not allowing" televised cigarette ads and pornography than of "censoring" them, and of "revenue enhancers" than of "taxes." In 2009, three in four Americans in one national survey approved of giving people "a choice" of public, government-run, or private health insurance. Yet in another survey, most Americans were not in favor of "creating a public health care plan administered by the federal government that would compete directly with private health insurance companies" (Stein, 2009). Because wording is such a delicate matter, critical thinkers will reflect on how the phrasing of a question might affect people's expressed opinions.

RANDOM SAMPLING

In everyday thinking, we tend to generalize from samples we observe, especially vivid cases. Given (a) a statistical summary of auto owners' evaluations of their car make and (b) the vivid comments of a biased sample—two frustrated owners—our impression may be influenced as much by the two unhappy owners as by the many more evaluations in the statistical summary. The temptation to ignore the **sampling bias** and to generalize from a few vivid but unrepresentative cases is nearly irresistible.

The point to remember: The best basis for generalizing is from a *representative sample.*

But it's not always possible to survey everyone in a group. So how do you obtain a representative sample—say, of the students at your high school? How could you choose a group that would represent the total student **population,** the whole group you want to study and describe? Typically, you would seek a **random sample,** in which every person in the entire group has an equal chance of participating. You might number the names in the general student listing and then use a random number generator to pick your survey participants. (Sending each student a questionnaire wouldn't work because the conscientious people who returned it would not be a random sample.) Large representative samples are better than small ones, but a small representative sample of 100 is better than an unrepresentative sample of 500.

Political pollsters sample voters in national election surveys just this way. Using only 1500 randomly sampled people, drawn from all areas of a country, they can provide a remarkably accurate snapshot of the nation's opinions. Without random sampling (also called *random selection*), large samples—including call-in phone samples and TV or website polls (think of *American Idol* fans voting)—often merely give misleading results.

The point to remember: Before accepting survey findings, think critically: Consider the sample. You cannot compensate for an unrepresentative sample by simply adding more people.

sampling bias a flawed sampling process that produces an unrepresentative sample.

population all those in a group being studied, from which samples may be drawn. (*Note:* Except for national studies, this does *not* refer to a country's whole population.)

random sample a sample that fairly represents a population because each member has an equal chance of inclusion.

FYI

With very large samples, estimates become quite reliable. *E* is estimated to represent 12.7 percent of the letters in written English. *E*, in fact, is 12.3 percent of the 925,141 letters in Melville's *Moby Dick*, 12.4 percent of the 586,747 letters in Dickens' *A Tale of Two Cities*, and 12.1 percent of the 3,901,021 letters in 12 of Mark Twain's works (*Chance News*, 1997).

Before You Move On

▶ ASK YOURSELF

Can you recall examples of misleading surveys you have experienced or read about? What survey principles did they violate?

▶ TEST YOURSELF

What are some strengths and weaknesses of the three different methods psychologists use to describe behavior—case studies, naturalistic observation, and surveys?

Answers to the Test Yourself questions can be found in Appendix E at the end of the book.

Module 5 Review

5-1 How do theories advance psychological science?

- Psychological *theories* are explanations that apply an integrated set of principles to organize observations and generate *hypotheses*—predictions that can be used to check the theory or produce practical applications of it. By testing their hypotheses, researchers can confirm, reject, or revise their theories.
- To enable other researchers to *replicate* the studies, researchers report them using precise *operational definitions* of their procedures and concepts. If others achieve similar results, confidence in the conclusion will be greater.

5-2 How do psychologists use case studies, naturalistic observation, and surveys to observe and describe behavior, and why is random sampling important?

- Description methods, which include *case studies, naturalistic observations,* and *surveys,* show us what can happen, and they may offer ideas for further study.
- The best basis for generalizing about a *population* is a representative sample; in a *random sample,* every person in the entire population being studied has an equal chance of participating.
- Descriptive methods describe but do not *explain* behavior, because these methods do not control for the many variables that can affect behavior.

Multiple-Choice Questions

1. Why is an operational definition necessary when reporting research findings?

a. An operational definition allows others to replicate the procedure.
b. An operational definition provides more context and includes many examples of the concept described.
c. An operational definition is easier to translate into multiple languages than a dictionary definition.
d. An operational definition uses more scientific language than a dictionary definition.
e. An operational definition is not necessary since a dictionary definition will work as well for replication.

2. A researcher looking for gender differences in 3-year-olds observes a preschool class and records how many minutes children of each gender play with dolls. She then compares the two sets of numbers. What type of descriptive research is she conducting?

a. Case study
b. National study
c. Random sample method
d. Naturalistic observation
e. Survey

3. Which of the following questions is best investigated by means of a survey?

a. Is IQ related to grades?
b. Are violent criminals genetically different from nonviolent criminals?
c. Does extra sleep improve memory?
d. What is the best study technique for AP® tests?
e. Are students more likely to be politically liberal or conservative?

4. A testable prediction that drives research is known as a(n)

a. theory.
b. hypothesis.
c. operational definition.
d. guess.
e. random sample.

5. Researchers are interested in finding out if winning Congressional candidates display more positive facial expressions than losing candidates. The researchers attend political debates and record how frequently each candidate displays positive facial expressions. Which research method are the researchers using?

a. Random sample
b. Case study
c. Naturalistic observation
d. Survey
e. Interview

6. An individual with an exceptional memory is identified. She is capable of recalling major events, the weather, and what she did on any given date. What research method is being used if a psychologist conducts an in-depth investigation of this individual including questionnaires, brain scans, and memory tests?

a. Naturalistic observation
b. Survey
c. Interview
d. Case study
e. Correlational method

7. Which of the following is most important when conducting survey research?

a. Choosing a representative sample
b. Choosing a large sample
c. Choosing a biased sample
d. Choosing a sample that includes every member of the population
e. Choosing a sample whose answers will likely support your hypothesis

Practice FRQs

1. A teacher wants to know if nightmares are more common than dreams. He asks volunteers from his second-period class to report how many dreams they had last week. He asks volunteers from his third-period class to report the number of nightmares they had last week. Describe two things wrong with the design of this study.

Answer (2 of the following)

1 point: There is no hypothesis stated.

1 point: In asking for volunteers, the teacher is taking a nonrandom sample that is probably not representative of the population of interest.

1 point: Neither dreams nor nightmares are operationally defined, so they might be interpreted differently by later researchers.

1 point: The research is not blind. The teacher could influence the results by the way he asked questions.

2. Naturalistic observation is a research method used by psychologists to investigate human and animal behavior. Identify three weaknesses of naturalistic observation.

(3 points)

Module 6

Correlation and Experimentation


imagebroker/Alamy

Module Learning Objectives

6-1 Describe positive and negative correlations, and explain how correlational measures can aid the process of prediction but not provide evidence of cause-effect relationships.

6-2 Explain illusory correlations.

6-3 Describe the characteristics of experimentation that make it possible to isolate cause and effect.

"Study finds that increased parental support for college results in lower grades" (Jaschik, 2013). "People with mental illness more likely to be smokers" (Belluck, 2013). What should we make of such news headlines—telling us that students whose parents pay the college bill tend to underachieve, and that smoking is associated with mental illness? Do these correlations indicate that students would achieve more if their parents became less supportive and that stopping smoking could produce better mental health? *No.* Read on.

Correlation

6-1 What are positive and negative correlations, and why do they enable prediction but not cause-effect explanation?

Describing behavior is a first step toward predicting it. Naturalistic observations and surveys often show us that one trait or behavior is related to another. In such cases, we say the two **correlate.** A statistical measure (the **correlation coefficient**) helps us figure how closely two things vary together, and thus how well either one predicts the other. Knowing how much aptitude test scores correlate with school success tells us how well the scores predict school success.

Throughout this book we will often ask how strongly two things are related: For example, how closely related are the personality scores of identical twins? How well do intelligence test scores predict career achievement? How closely is stress related to disease? In such cases, **scatterplots** can be very revealing.

Each dot in a scatterplot represents the values of two variables. The three scatterplots in **FIGURE 6.1** illustrate the range of possible correlations from a perfect positive to a perfect negative. (Perfect correlations rarely occur in the "real world.") A correlation is positive if two sets of scores, such as height and weight, tend to rise or fall together.

correlation a measure of the extent to which two variables change together, and thus of how well either variable predicts the other.

correlation coefficient a statistical index of the relationship between two variables (from –1.0 to +1.0).

scatterplot a graphed cluster of dots, each of which represents the values of two variables. The slope of the points suggests the direction of the relationship between the two variables. The amount of scatter suggests the strength of the correlation (little scatter indicates high correlation).

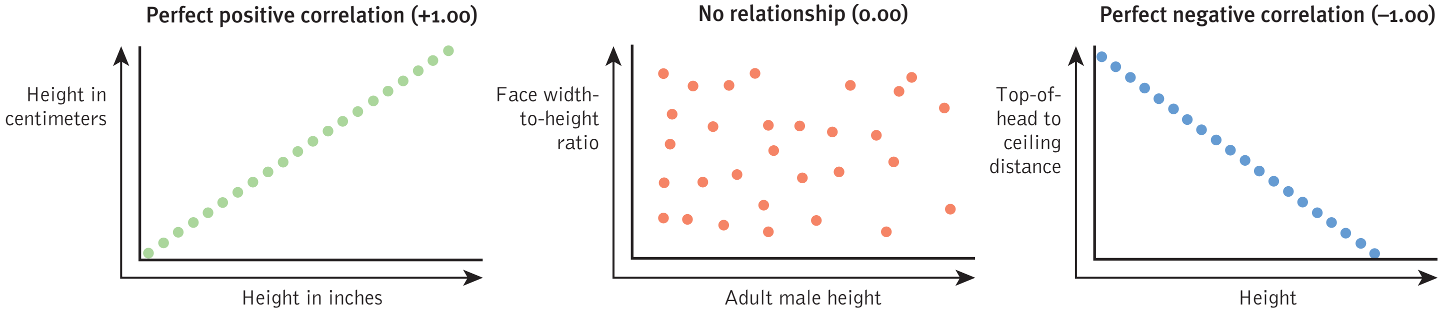


Figure 6.1

Scatterplots, showing patterns of correlation Correlations can range from +1.00 (scores on one measure increase in direct proportion to scores on another) to –1.00 (scores on one measure decrease precisely as scores rise on the other).

Saying that a correlation is "negative" says nothing about its strength or weakness. A correlation is negative if two sets of scores relate inversely, one set going up as the other goes down. The study of Nevada university students' inner speech discussed in Module 5 also included a correlational component. Students' reports of inner speech correlated negatively (–.36) with their scores on another measure: psychological distress. Those who reported more inner speech tended to report slightly *less* psychological distress.

Statistics can help us see what the naked eye sometimes misses. To demonstrate this for yourself, try an imaginary project. Wondering if tall men are more or less easygoing, you collect two sets of scores: men's heights and men's temperaments. You measure the heights of 20 men, and you have someone else independently assess their temperaments (from zero for extremely calm to 100 for highly reactive).

With all the relevant data right in front of you (**TABLE 6.1**), can you tell whether the correlation between height and reactive temperament is positive, negative, or close to zero?

Comparing the columns in Table 6.1, most people detect very little relationship between height and temperament. In fact, the correlation in this imaginary example is positive, +0.63, as we can see if we display the data as a scatterplot. In **FIGURE 6.2** on the next page, moving from left to right, the upward, oval-shaped slope of the cluster of points shows that our two imaginary sets of scores (height and temperament) tend to rise together.

If we fail to see a relationship when data are presented as systematically as in Table 6.1, how much less likely are we to notice them in everyday life? To see what is right in front of us, we sometimes need statistical illumination. We can easily see evidence of gender discrimination when given statistically summarized information about job level, seniority, performance,

Table 6.1 Height and Temperamental Reactivity of 20 Men

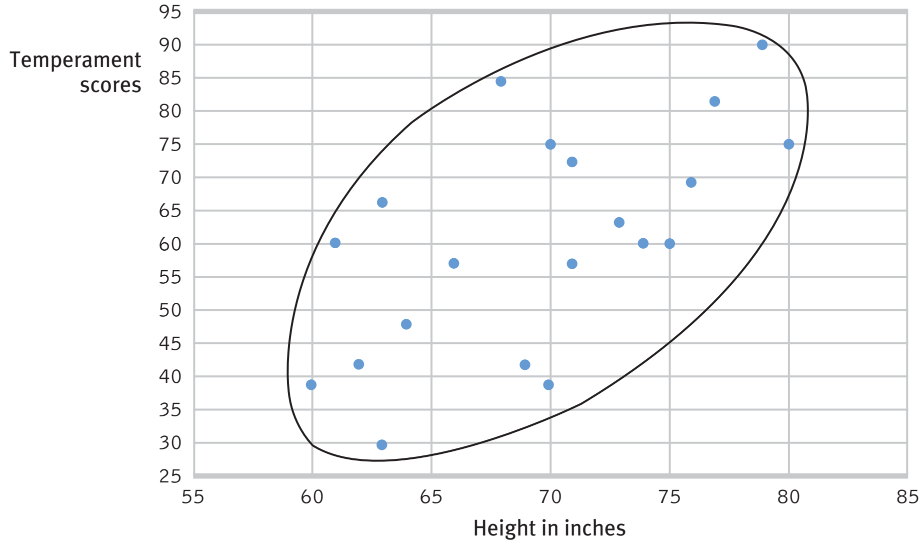
Person	Height in Inches	Temperament
1	80	75
2	63	66
3	61	60
4	79	90
5	74	60
6	69	42
7	62	42
8	75	60
9	77	81
10	60	39
11	64	48
12	76	69
13	71	72
14	66	57
15	73	63
16	70	75
17	63	30
18	71	57
19	68	84
20	70	39

AP® Exam Tip

This is the first of several times in your psychology course that you will see something labeled as being positive or negative. We often think that if something is positive it is good and if it's negative it's bad. That is rarely the case in this course. Here, positive and negative refer only to the direction of the correlation. They say nothing about whether the relationship is desirable or not.

Figure 6.2

Scatterplot for height and reactive temperament This display of data from 20 imagined people (each represented by a data point) reveals an upward slope, indicating a positive correlation. The considerable scatter of the data indicates the correlation is much lower than +1.0.



gender, and salary. But we often see no discrimination when the same information dribbles in, case by case (Twiss et al., 1989). See **TABLE 6.2** to test your understanding further.

The point to remember: A correlation coefficient, which can range from –1.0 to +1.0, reveals the extent to which two things relate. The closer the score gets to –1 or +1, the stronger the correlation.

Table 6.2
Test your understanding of correlation. Which of the following news reports are examples of a *positive* correlation, and which are examples of a *negative* correlation? (Check your answers below.)
1. The more children and youth used various media, the less happy they were with their lives (Kaiser, 2010). ___________
2. The less sexual content teens saw on TV, the less likely they were to have sex (Collins et al., 2004). ___________
3. The longer children were breast-fed, the greater their later academic achievement (Horwood & Ferguson, 1998). ___________
4. The more income rose among a sample of poor families, the fewer psychiatric symptoms their children experienced (Costello et al., 2003). ___________
ANSWERS: 1. negative, 2. positive, 3. positive, 4. negative

Correlation and Causation

Correlations help us predict. The *New York Times* reports that U.S. counties with high gun ownership rates tend to have high murder rates (Luo, 2011). Gun ownership predicts homicide. What might explain this guns-homicide correlation?

I can almost hear someone thinking, "Well, of course, guns kill people, often in moments of passion." If so, that could be an example of A (guns) causes B (murder). But I can hear other readers saying, "Not so fast. Maybe people in dangerous places buy more guns for self-protection—maybe B causes A." Or maybe some third variable C causes both A and B.

© Nancy Brown/Getty Images

Correlation need not mean causation Length of marriage correlates with hair loss in men. Does this mean that marriage *causes* men to lose their hair (or that balding men make better husbands)? In this case, as in many others, a third variable probably explains the correlation: Golden anniversaries and baldness both accompany aging.

Another example: Self-esteem correlates negatively with (and therefore predicts) depression. (The lower people's self-esteem, the more they are at risk for depression.) So, does low self-esteem *cause* depression? If, based on the correlational evidence, you assume that it does, you have much company. A nearly irresistible thinking error is assuming that an association, sometimes presented as a correlation coefficient, proves causation. But no matter how strong the relationship, it does not.

As options 2 and 3 in **FIGURE 6.3** show, we'd get the same negative correlation between self-esteem and depression if depression caused people to be down on themselves, or if some third variable—such as heredity or brain chemistry—caused both low self-esteem and depression.

This point is so important—so basic to thinking smarter with psychology—that it merits one more example. A survey of over 12,000 adolescents found that the more teens feel loved by their parents, the less likely they are to behave in unhealthy ways—having early sex, smoking, abusing alcohol and drugs, exhibiting violence (Resnick et al., 1997). "Adults have a powerful effect on their children's behavior right through the high school years," gushed an Associated Press (AP) story reporting the finding. But this correlation comes with no built-in cause-effect arrow. The AP could as well have reported, "Well-behaved teens feel their parents' love and approval; out-of-bounds teens more often think their parents are disapproving jerks."

The point to remember (turn the volume up here): *Association does not prove causation.*[1] Correlation indicates the *possibility* of a cause-effect relationship *but does not prove such.* Remember this principle and you will be wiser as you read and hear news of scientific studies.

FYI

A *New York Times* writer reported a massive survey showing that "adolescents whose parents smoked were 50 percent more likely than children of nonsmokers to report having had sex." He concluded (would you agree?) that the survey indicated a causal effect—that "to reduce the chances that their children will become sexually active at an early age" parents might "quit smoking" (O'Neil, 2002).

AP® Exam Tip

Take note of how much emphasis is put on this idea. Correlation and association do not prove a cause-effect relationship.

[1]Because many associations are stated as correlations, the famously worded principle is "Correlation does not prove causation." That's true, but it's also true of associations verified by other nonexperimental statistics (Hatfield et al., 2006).

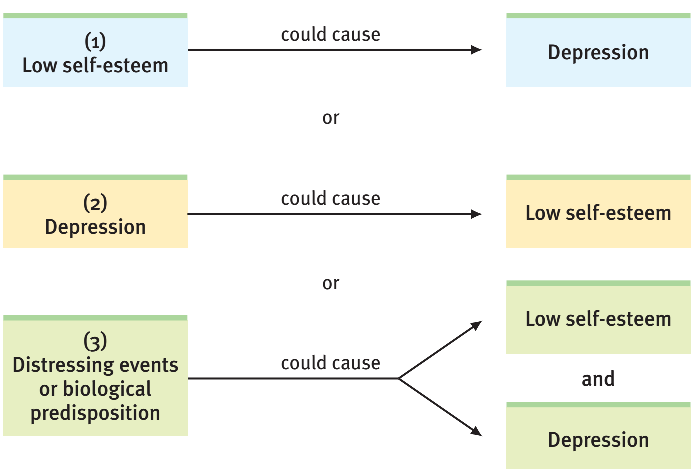


Figure 6.3

Three possible cause-effect relationships People low in self-esteem are more likely to report depression than are those high in self-esteem. One possible explanation of this negative correlation is that a bad self-image causes depressed feelings. But, as the diagram indicates, other cause-effect relationships are possible.

Illusory Correlations

6-2 What are illusory correlations?

Correlation coefficients make visible the relationships we might otherwise miss. They also restrain our "seeing" relationships that actually do not exist. A perceived but nonexistent correlation is an **illusory correlation.** When we *believe* there is a relationship between two things, we are likely to *notice* and *recall* instances that confirm our belief (Trolier & Hamilton, 1986).

illusory correlation the perception of a relationship where none exists.

Because we are sensitive to dramatic or unusual events, we are especially likely to notice and remember the occurrence of two such events in sequence—say, a premonition of an unlikely phone call followed by the call. When the call does not follow the premonition, we are less likely to note and remember the nonevent. Illusory correlations help explain many superstitious beliefs, such as the presumption that infertile couples who adopt become more likely to conceive (Gilovich, 1991). Couples who conceive after adopting capture our attention. We're less likely to notice those who adopt and never conceive, or those who conceive without adopting. In other words, illusory correlations occur when we over-rely on the top left cell of **FIGURE 6.4**, ignoring equally essential information in the other cells.

Such illusory thinking helps explain why for so many years people believed (and many still do) that sugar makes children hyperactive, that getting chilled and wet causes people to catch a cold, and that changes in the weather trigger arthritis pain. We are, it seems, prone to perceiving patterns, whether they're there or not.

The point to remember: When we notice random coincidences, we may forget that they are random and instead see them as correlated. Thus, we can easily deceive ourselves by seeing what is not there.

FYI

A study reported in the *British Medical Journal* found that youths who identify with the goth subculture attempt, more often than other young people, to harm or kill themselves (Young et al., 2006). Can you imagine multiple possible explanations for this association?

Figure 6.4

Illusory correlation in everyday life Many people believe infertile couples become more likely to conceive a child after adopting a baby. This belief arises from their attention being drawn to such cases. The many couples who adopt without conceiving or conceive without adopting grab less attention. To determine whether there actually is a correlation between adoption and conception, we need data from all four cells in this figure. (From Gilovich, 1991.)

	Conceive	Do not conceive
Adopt	confirming evidence	disconfirming evidence
Do not adopt	disconfirming evidence	confirming evidence

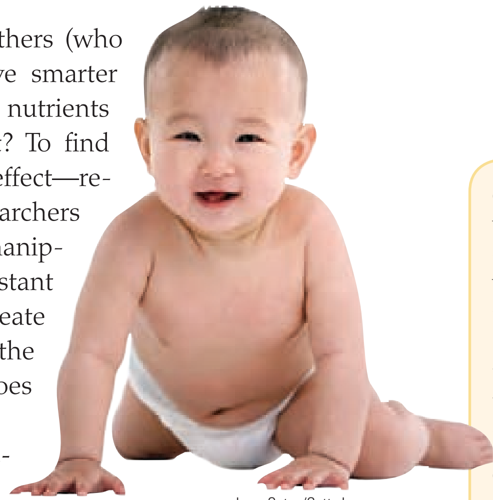
Michael Newman Jr./PhotoEdit

Experimentation

6-3 What are the characteristics of experimentation that make it possible to isolate cause and effect?

Happy are they, remarked the Roman poet Virgil, "who have been able to perceive the causes of things." How might psychologists perceive causes in correlational studies, such as the correlation between breast feeding and intelligence?

Researchers have found that the intelligence scores of children who were breast-fed as infants are somewhat higher than the scores of children who were bottle-fed (Angelsen et al., 2001; Mortensen et al., 2002; Quinn et al., 2001). In Britain, breast-fed babies have also been more likely than their bottle-fed counterparts to eventually move into a higher social class (Martin et al., 2007). The "breast is best" intelligence effect shrinks when researchers compare breast-fed and bottle-fed children from the same families (Der et al., 2006).


Lane Oatey/Getty Images

What do such findings mean? Do smarter mothers (who in modern countries more often breast feed) have smarter children? Or, as some researchers believe, do the nutrients of mother's milk contribute to brain development? To find answers to such questions—to isolate cause and effect—researchers can **experiment.** Experiments enable researchers to isolate the effects of one or more variables by (1) manipulating the variables of interest and (2) holding constant ("controlling") other variables. To do so, they often create an **experimental group,** in which people receive the treatment, and a contrasting **control group** that does not receive the treatment.

Earlier we mentioned the place of *random sampling* in a well-done survey. Consider now the equally important place of *random assignment* in a well-done experiment. To minimize any preexisting differences between the two groups, researchers **randomly assign** people to the two conditions. Random assignment effectively equalizes the two groups. If one-third of the volunteers for an experiment can wiggle their ears, then about one-third of the people in each group will be ear wigglers. So, too, with ages, attitudes, and other characteristics, which will be similar in the experimental and control groups. Thus, if the groups differ at the experiment's end, we can surmise that the treatment had an effect.

To experiment with breast feeding, one research team randomly assigned some 17,000 Belarus newborns and their mothers either to a breast-feeding promotion group or to a normal pediatric care program (Kramer et al., 2008). At 3 months of age, 43 percent of the infants in the experimental group were being exclusively breast-fed, as were 6 percent in the control group. At age 6, when nearly 14,000 of the children were restudied, those who had been in the breast-feeding promotion group had intelligence test scores averaging six points higher than their control condition counterparts.

No single experiment is conclusive, of course. But randomly assigning participants to one feeding group or the other effectively eliminated all variables except nutrition. This supported the conclusion that breast is indeed best for developing intelligence: If a behavior (such as test performance) changes when we vary an experimental variable (such as infant nutrition), then we infer the variable is having an effect.

The point to remember: Unlike correlational studies, which uncover naturally occurring relationships, an experiment manipulates a variable to determine its effect.

Consider, then, how we might assess therapeutic interventions. Our tendency to seek new remedies when we are ill or emotionally down can produce misleading testimonies. If three days into a cold we start taking vitamin C tablets and find our cold symptoms lessening, we may credit the pills rather than the cold naturally subsiding. In the 1700s, bloodletting seemed effective. People sometimes improved after the treatment; when they didn't, the practitioner inferred the disease was too advanced to be reversed. So, whether or not a remedy is truly effective, enthusiastic users will probably endorse it. To determine its effect, we must control for other variables.

And that is precisely how investigators evaluate new drug treatments and new methods of psychological therapy (Modules 72–73). They randomly assign participants in these studies to research groups. One group receives a treatment (such as a medication). The other group receives a pseudotreatment—an inert placebo (perhaps a pill with no drug in it). The participants are often blind (uninformed) about what treatment, if any, they are receiving. If the study is using a **double-blind procedure,** neither the participants nor the research assistants who administer the drug and collect the data will know which group is receiving the treatment.

experiment a research method in which an investigator manipulates one or more factors (independent variables) to observe the effect on some behavior or mental process (the dependent variable). By *random assignment* of participants, the experimenter aims to control other relevant variables.

experimental group in an experiment, the group exposed to the treatment, that is, to one version of the independent variable.

control group in an experiment, the group *not* exposed to the treatment; contrasts with the experimental group and serves as a comparison for evaluating the effect of the treatment.

random assignment assigning participants to experimental and control groups by chance, thus minimizing preexisting differences between the different groups.

double-blind procedure an experimental procedure in which both the research participants and the research staff are ignorant (blind) about whether the research participants have received the treatment or a placebo. Commonly used in drug-evaluation studies.

DORA: removed acronyms from AP Exam Tip--per note from Christine. --JO'N

In such studies, researchers can check a treatment's actual effects apart from the participants' and the staff's belief in its healing powers. Just thinking you are getting a treatment can boost your spirits, relax your body, and relieve your symptoms. This **placebo effect** is well documented in reducing pain, depression, and anxiety (Kirsch, 2010). And the more expensive the placebo, the more "real" it seems to us—a fake pill that costs $2.50 works better than one costing 10 cents (Waber et al., 2008). To know how effective a therapy really is, researchers must control for a possible placebo effect.

placebo [pluh-SEE-bo; Latin for "I shall please"] **effect** experimental results caused by expectations alone; any effect on behavior caused by the administration of an inert substance or condition, which the recipient assumes is an active agent.

independent variable the experimental factor that is manipulated; the variable whose effect is being studied.

confounding variable a factor other than the independent variable that might produce an effect in an experiment.

dependent variable the outcome factor; the variable that may change in response to manipulations of the independent variable.

Independent and Dependent Variables

Here is a practical experiment: In a not yet published study, Victor Benassi and his colleagues gave college psychology students frequent in-class quizzes. Some items served merely as *review*—students were given questions with answers. Other *self-testing* items required students to actively produce the answers. When tested weeks later on a final exam, students did far better on material on which they had been tested (75 percent correct) rather than merely reviewed (51 percent correct). By a wide margin, testing beat restudy.

This simple experiment manipulated just one factor: the study procedure (reading answers versus self-testing). We call this experimental factor the **independent variable** because we can vary it *independently* of other factors, such as the students' memories, intelligence, and age. These other factors, which can potentially influence the results of the experiment, are called **confounding variables.** Random assignment controls for possible confounding variables.

Experiments examine the effect of one or more independent variables on some measurable behavior, called the **dependent variable** because it can vary *depending* on what takes place during the experiment. Both variables are given precise *operational definitions,* which specify the procedures that manipulate the independent variable (the review versus self-testing study method in this analysis) or measure the dependent variable (final exam performance). These definitions answer the "What do you mean?" question with a level of precision that enables others to repeat the study. (See **FIGURE 6.5** for the previously mentioned breast-milk experiment's design.)

Let's pause to check your understanding using a simple psychology experiment: To test the effect of perceived ethnicity on the availability of a rental house, researchers sent identically worded e-mail inquiries to 1115 Los Angeles-area landlords (Carpusor & Loges, 2006). The researchers varied the ethnic connotation of the sender's name and tracked the percentage of positive replies (invitations to view the apartment in person). "Patrick McDougall," "Said Al-Rahman," and "Tyrell Jackson" received, respectively, 89 percent, 66 percent, and 56 percent invitations. (In this experiment, what was the independent variable? The dependent variable?[2])

AP® Exam Tip

The identification of independent and dependent variables is the single most likely concept to be tested on the AP® exam. Experiments are critical to psychology, and independent and dependent variables are critical to experiments.

Figure 6.5
Experimentation To discern causation, psychologists may randomly assign some participants to an experimental group, others to a control group. Measuring the dependent variable (intelligence score in later childhood) will determine the effect of the independent variable (whether breast feeding was promoted).

Random assignment (controlling for other confounding variables, such as parental intelligence and environment)

© Radius Images/Alamy

Group	Independent variable	Dependent variable
Experimental	Promoted breast feeding	Intelligence score, age 6
Control	Did not promote breast feeding	Intelligence score, age 6

A key goal of experimental design is **validity,** which means the experiment will test what it is supposed to test. In the rental housing experiment, we might ask, "Did the e-mail inquiries test the effect of perceived ethnicity? Did the landlords' response actually vary with the ethnicity of the name?"

validity the extent to which a test or experiment measures or predicts what it is supposed to.

Experiments can also help us evaluate social programs. Do early childhood education programs boost impoverished children's chances for success? What are the effects of different antismoking campaigns? Do school sex-education programs reduce teen pregnancies? To answer such questions, we can experiment: If an intervention is welcomed but resources are scarce, we could use a lottery to randomly assign some people (or regions) to experience the new program and others to a control condition. If later the two groups differ, the intervention's effect will be supported (Passell, 1993).

Let's recap. A *variable* is anything that can vary (infant nutrition, intelligence, TV exposure—anything within the bounds of what is feasible and ethical). Experiments aim to *manipulate* an *independent* variable, *measure* the *dependent* variable, and allow random assignment to *control* all other variables. An experiment has at least two different conditions: an *experimental condition* and a *comparison* or *control condition. Random assignment* works to equate the groups before any treatment effects occur. In this way, an experiment tests the effect of at least one independent variable (what we manipulate) on at least one dependent variable (the outcome we measure). **TABLE 6.3** compares the features of psychology's research methods.

AP® Exam Tip

Almost 15 pages of text are summarized in this one table. Spend some time with it, as it is information you will likely encounter on the AP® exam.

[2]The independent variable, which the researchers manipulated, was the ethnicity-related names. The dependent variable, which they measured, was the positive response rate.

Table 6.3 Comparing Research Methods

Research Method	Basic Purpose	How Conducted	What Is Manipulated	Strengths	Weaknesses
Descriptive	To observe and record behavior	Do case studies, naturalistic observations, or surveys	Nothing	Case studies require only one participant; naturalistic observations may be done when it is not ethical to manipulate variables; surveys may be done quickly and inexpensively (compared with experiments)	Uncontrolled variables mean cause and effect cannot be determined; single cases may be misleading
Correlational	To detect naturally occurring relationships; to assess how well one variable predicts another	Collect data on two or more variables; no manipulation	Nothing	Works with large groups of data, and may be used in situations where an experiment would not be ethical or possible	Does not specify cause and effect
Experimental	To explore cause and effect	Manipulate one or more variables; use random assignment	The independent variable(s)	Specifies cause and effect, and variables are controlled	Sometimes not feasible; results may not generalize to other contexts; not ethical to manipulate certain variables

"If I don't think it's going to work, will it still work?"

Before You Move On

▶ ASK YOURSELF

If you were to become a research psychologist, what questions would you like to explore with experiments?

▶ TEST YOURSELF

Why, when testing a new drug to control blood pressure, would we learn more about its effectiveness from giving it to half of the participants in a group of 1000 than to all 1000 participants?

Answers to the Test Yourself questions can be found in Appendix E at the end of the book.

Module 6 Review

6-1 What are positive and negative correlations, and why do they enable prediction but not cause-effect explanation?

- In a positive correlation, two variables rise or fall together. In a negative correlation, one item rises as the other falls.
- *Scatterplots* can help us to see *correlations*.
- A *correlation coefficient* can describe the strength and direction of a relationship between two variables, from +1.00 (a perfect positive correlation) through zero (no correlation at all) to –1.00 (a perfect negative correlation).

6-2 What are illusory correlations?

- *Illusory correlations* are random events that we notice and falsely assume are related.
- Patterns or sequences occur naturally in sets of random data, but we tend to interpret these patterns as meaningful connections, perhaps in an attempt to make sense of the world around us.

6-3 What are the characteristics of experimentation that make it possible to isolate cause and effect?

- To discover cause-effect relationships, psychologists conduct *experiments,* manipulating one or more variables of interest and controlling other variables.
- Using *random assignment,* they can minimize *confounding variables,* such as preexisting differences between the *experimental group* (exposed to the treatment) and the *control group* (given a placebo or different version of the treatment).
- The *independent variable* is the factor the experimenter manipulates to study its effect; the *dependent variable* is the factor the experimenter measures to discover any changes occurring in response to the manipulation of the independent variable.
- Studies may use a *double-blind procedure* to avoid the *placebo effect* and researcher's bias.
- An experiment has *validity* if it tests what it is supposed to test.

Multiple-Choice Questions

1. Which of the following is an example of negative correlation?

a. People who spend more time exercising tend to weigh less.
b. Teenage females tend to have fewer speeding tickets than teenage males.
c. Students with low IQ scores tend to have lower grades.
d. As hours studying for a test decrease, so do grades on that test.
e. Students' shoe sizes are not related to their grades.

2. Which of the following is used only in correlation studies?

a. Double blind
b. Placebo
c. Random assignment
d. Scatterplot
e. Random sample

3. Researchers have discovered that individuals with lower income levels report having fewer hours of total sleep. Therefore,

a. income and sleep levels are positively correlated.
b. income and sleep levels are negatively correlated.
c. income and sleep levels are inversely correlated.
d. income and sleep levels are not correlated.
e. lower income levels cause individuals to have fewer hours of sleep.

4. Which of the following correlation coefficients represents the strongest relationship between two variables?

a. +.30
b. +.75
c. +1.3
d. –.85
e. –1.2

5. The purpose of random assignment is to

a. allow participants in both the experimental and control groups to be exposed to the independent variable.
b. ensure that every member of the population had an equal chance of being selected to participate in the research.
c. eliminate the placebo effect.
d. reduce potential confounding variables.
e. generate operational definitions for the independent and dependent variables.

Practice FRQs

1. Students with higher scores on anxiety scales were found to have lower scores on standardized tests. What research method would show this relationship? Why can no cause-effect conclusion be drawn from the results?

Answer

1 point: This research method is a correlation study.

1 point: There are three possibilities for causation: Anxiety could cause low test scores, low test scores could cause anxiety, or a third factor could cause both anxiety and low test scores. No conclusions can be drawn about causation because this is not an experiment.

2. Ms. Ledbetter wants to determine if the new review activity she developed will improve student performance on unit exams. She randomly separates 160 students into two groups. Group A reviews for the unit exam in the traditional manner they have always used. Group B participates in the new review activity. After reviewing, both groups are given the same unit exam and their scores are compared. Identify the independent and dependent variables for this experiment.

(2 points)

Module 7

Statistical Reasoning in Everyday Life


Norm Hall/Getty Images

Module Learning Objectives

7-1 Describe the three measures of central tendency, and discuss the relative usefulness of the two measures of variation.

7-2 Explain how we know whether an observed difference can be generalized to other populations.

In descriptive, correlational, and experimental research, statistics are tools that help us see and interpret what the unaided eye might miss. Sometimes the unaided eye misses badly. Researchers invited 5522 Americans to estimate the percentage of wealth possessed by the richest 20 percent in their country (Norton & Ariely, 2011). Their average person's guess—58 percent—"dramatically underestimated" the actual wealth inequality. (The wealthiest 20 percent possess 84 percent of the wealth.)

FYI

Asked about the *ideal* wealth distribution in America, Democrats and Republicans were surprisingly similar. In the Democrats' ideal world, the richest 20 percent would possess 30 percent of the wealth. The Republicans' ideal world was similar, with the richest 20 percent possessing 35 percent of the wealth. (Norton & Ariely, 2011).

The Need for Statistics

Accurate statistical understanding benefits everyone. To be an educated person today is to be able to apply simple statistical principles to everyday reasoning. One needn't memorize complicated formulas to think more clearly and critically about data.

Off-the-top-of-the-head estimates often misread reality and then mislead the public. Someone throws out a big, round number. Others echo it, and before long the big, round number becomes public misinformation. A few examples:

- *Ten percent of people are lesbians or gay men.* Or is it 2 to 3 percent, as suggested by various national surveys (Module 53)?
- *We ordinarily use but 10 percent of our brain.* Or is it closer to 100 percent (Module 12)?
- *The human brain has 100 billion nerve cells.* Or is it more like 40 billion, as suggested by extrapolation from sample counts (Module 10)?

The point to remember: Doubt big, round, undocumented numbers.

Patrick Hardin/cartoonstock

"Figures can be misleading—so I've written a song which I think expresses the real story of the firm's performance this quarter."

AP® Exam Tip

Do math and statistics scare you? Take a couple of deep breaths and relax before continuing. You will not be asked to do difficult computations on the AP® exam. Nothing will be beyond the scope of simple mental math. You need to focus on the concepts. Why do these statistics exist? How can they help us understand the real world?

Statistical illiteracy also feeds needless health scares (Gigerenzer et al., 2008, 2009, 2010). In the 1990s, the British press reported a study showing that women taking a particular contraceptive pill had a 100 percent increased risk of blood clots that could produce strokes. This caused thousands of women to stop taking the pill, leading to a wave of unwanted pregnancies and an estimated 13,000 additional abortions (which also are associated with increased blood clot risk). And what did the study find? A 100 percent increased risk, indeed—but only from 1 in 7000 women to 2 in 7000 women. Such false alarms underscore the need to teach statistical reasoning and to present statistical information more transparently.

Descriptive Statistics

7-1 How do we describe data using three measures of central tendency, and what is the relative usefulness of the two measures of variation?

Once researchers have gathered their data, they may use **descriptive statistics** to organize that data meaningfully. One way to do this is to convert the data into a simple *bar graph,* called a **histogram,** as in **FIGURE 7.1,** which displays a distribution of different brands of trucks still on the road after a decade. When reading statistical graphs such as this, take care. It's easy to design a graph to make a difference look big (Figure 7.1a) or small (Figure 7.1b). The secret lies in how you label the vertical scale (the *y-axis*).

The point to remember: Think smart. When viewing figures in magazines and on television, read the scale labels and note their range.

Measures of Central Tendency

The next step is to summarize the data using some *measure of central tendency,* a single score that represents a whole set of scores. The simplest measure is the **mode,** the most frequently occurring score or scores. The most commonly reported is the **mean,** or arithmetic average—the total sum of all the scores divided by the number of scores. On a divided highway, the median is the middle. So, too, with data: The **median** is the midpoint—the 50th percentile. If you arrange all the scores in order from the highest to the lowest, half will be above the median and half will be below it. In a symmetrical, bell-shaped distribution of scores, the mode, mean, and median scores may be the same or very similar.

descriptive statistics numerical data used to measure and describe characteristics of groups. Includes measures of central tendency and measures of variation.

histogram a bar graph depicting a frequency distribution.

mode the most frequently occurring score(s) in a distribution.

mean the arithmetic average of a distribution, obtained by adding the scores and then dividing by the number of scores.

median the middle score in a distribution; half the scores are above it and half are below it.

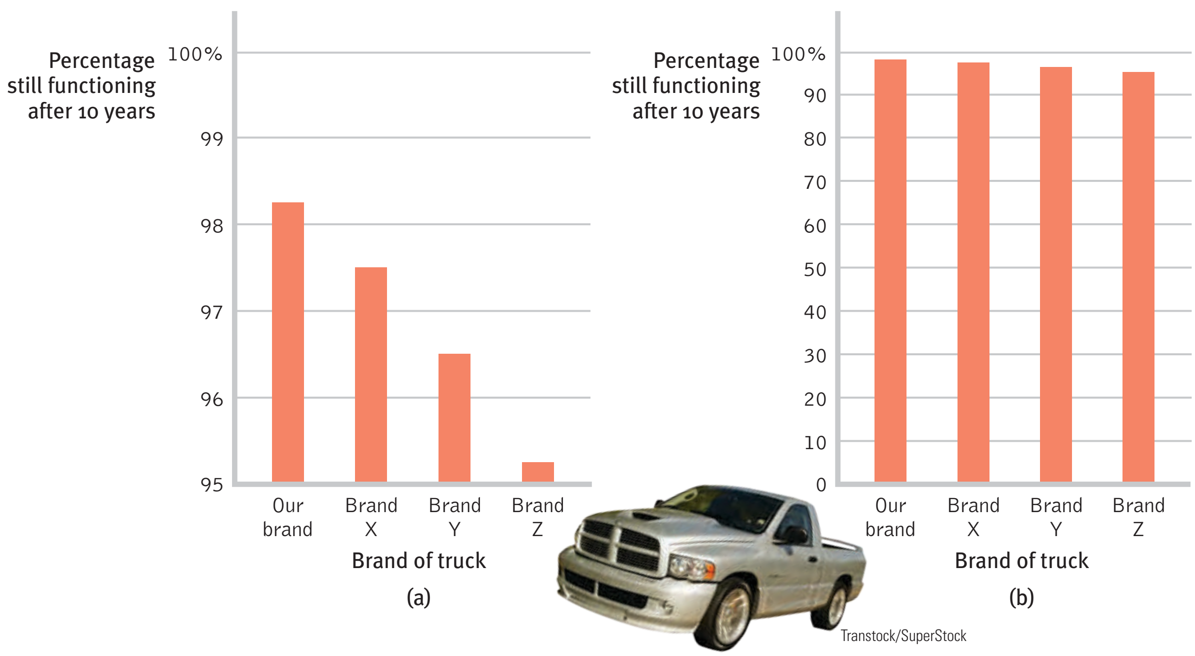


Figure 7.1 **Read the scale labels** An American truck manufacturer offered graph (a)—with actual brand names included—to suggest the much greater durability of its trucks. Note, however, how the apparent difference shrinks as the vertical scale changes in graph (b).

Measures of central tendency neatly summarize data. But consider what happens to the mean when a distribution is lopsided, or **skewed,** by a few way-out scores. With income data, for example, the mode, median, and mean often tell very different stories (**FIGURE 7.2**). This happens because the mean is biased by a few extreme scores. When Microsoft co-founder Bill Gates sits down in an intimate café, its average (mean) customer instantly becomes a billionaire. But the customers' median wealth remains unchanged. Understanding this, you can see how a British newspaper could accurately run the headline "Income for 62% Is Below Average" (Waterhouse, 1993). Because the bottom *half* of British income earners receive only a *quarter* of the national income cake, most British people, like most people everywhere, make less than the mean. Mean and median tell different true stories.

FYI
The average person has one ovary and one testicle.

The point to remember: Always note which measure of central tendency is reported. If it is a mean, consider whether a few atypical scores could be distorting it.

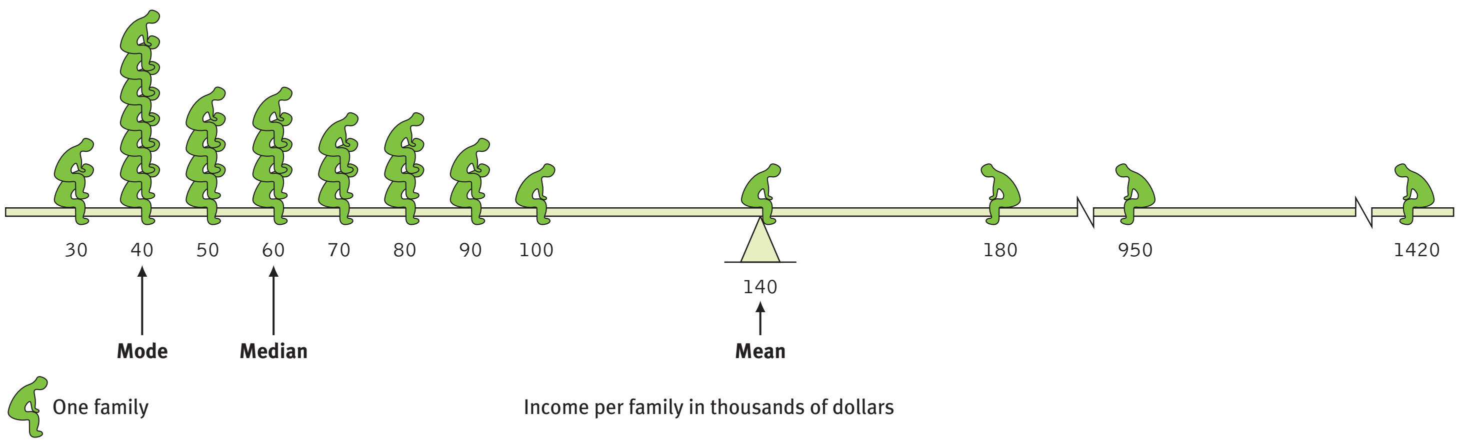


Figure 7.2
A skewed distribution This graphic representation of the distribution of a village's incomes illustrates the three measures of central tendency—mode, median, and mean. Note how just a few high incomes make the mean—the fulcrum point that balances the incomes above and below—deceptively high.

Measures of Variation

Knowing the value of an appropriate measure of central tendency can tell us a great deal. But the single number omits other information. It helps to know something about the amount of *variation* in the data—how similar or diverse the scores are. Averages derived from scores with low variability are more reliable than averages based on scores with high variability. Consider a basketball player who scored between 13 and 17 points in each of her first 10 games in a season. Knowing this, we would be more confident that she would score near 15 points in her next game than if her scores had varied from 5 to 25 points.

The **range** of scores—the gap between the lowest and highest scores—provides only a crude estimate of variation. A couple of extreme scores in an otherwise uniform group, such as the $950,000 and $1,420,000 incomes in Figure 7.2, will create a deceptively large range.

The more useful standard for measuring how much scores deviate from one another is the **standard deviation.** It better gauges whether scores are packed together or dispersed, because it uses information from each score (**TABLE 7.1**). The computation assembles information about how much individual scores differ from the mean. If your high school serves a community where most families have similar incomes, family income data will have a relatively small standard deviation compared with the more diverse community population outside your school.

You can grasp the meaning of the standard deviation if you consider how scores tend to be distributed in nature. Large numbers of data—heights, weights, intelligence scores, grades (though not incomes)—often form a symmetrical, *bell-shaped* distribution.

skewed distribution a representation of scores that lack symmetry around their average value.

range the difference between the highest and lowest scores in a distribution.

standard deviation a computed measure of how much scores vary around the mean score.

Table 7.1 Standard Deviation Is Much More Informative Than Mean Alone

Note that the test scores in Class A and Class B have the same mean (80), but very different standard deviations, which tell us more about how the students in each class are really faring.

Test Scores in Class A			Test Scores in Class B		
Score	Deviation from the Mean	Squared Deviation	Score	Deviation from the Mean	Squared Deviation
72	–8	64	60	–20	400
74	–6	36	60	–20	400
77	–3	9	70	–10	100
79	–1	1	70	–10	100
82	+2	4	90	+10	100
84	+4	16	90	+10	100
85	+5	25	100	+20	400
87	+7	49	100	+20	400
Total = 640	Sum of (deviations)^2 = 204		Total = 640	Sum of (deviations)^2 = 2000	
Mean = 640 ÷ 8 = 80			**Mean** = 640 ÷ 8 = 80		

Standard deviation =

$$\sqrt{\frac{\text{Sum of (deviations)}^2}{\text{Number of scores}}} = \sqrt{\frac{204}{8}} = 5.0$$

Standard deviation =

$$\sqrt{\frac{\text{Sum of (deviations)}^2}{\text{Number of scores}}} = \sqrt{\frac{2000}{8}} = 15.8$$

Most cases fall near the mean, and fewer cases fall near either extreme. This bell-shaped distribution is so typical that we call the curve it forms the **normal curve.**

As **FIGURE 7.3** shows, a useful property of the normal curve is that roughly 68 percent of the cases fall within one standard deviation on either side of the mean. About 95 percent of cases fall within two standard deviations. Thus, as Module 61 notes, about 68 percent of people taking an intelligence test will score within ±15 points of 100. About 95 percent will score within ±30 points.

normal curve *(normal distribution)* a symmetrical, bell-shaped curve that describes the distribution of many types of data; most scores fall near the mean (about 68 percent fall within one standard deviation of it) and fewer and fewer near the extremes.

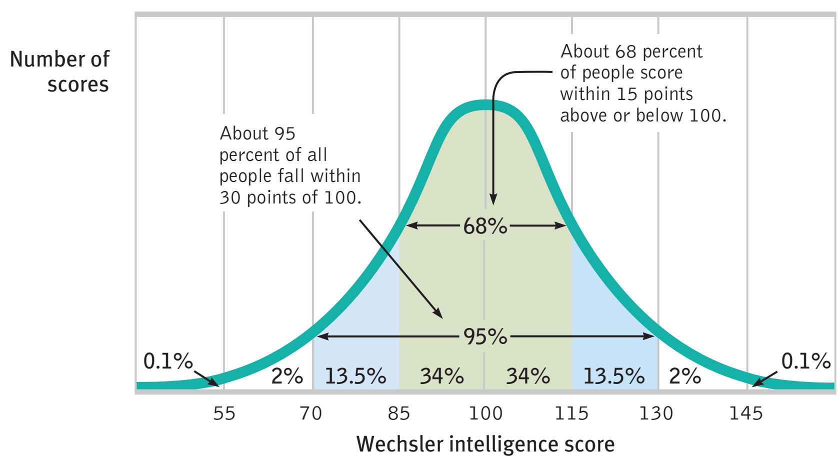


Figure 7.3
The normal curve Scores on aptitude tests tend to form a normal, or bell-shaped, curve. For example, the most commonly used intelligence test, the Wechsler Adult Intelligence Scale, calls the average score 100.

Inferential Statistics

7-2 How do we know whether an observed difference can be generalized to other populations?

Data are "noisy." The average score in one group (breast-fed babies) could conceivably differ from the average score in another group (bottle-fed babies) not because of any real difference but merely because of chance fluctuations in the people sampled. How confidently, then, can we infer that an observed difference is not just a fluke—a chance result of your sampling? For guidance, we can ask how reliable and significant the differences are. These **inferential statistics** help us determine if results can be generalized to a larger population.

inferential statistics numerical data that allow one to generalize—to infer from sample data the probability of something being true of a population.

When Is an Observed Difference Reliable?

In deciding when it is safe to generalize from a sample, we should keep three principles in mind.

1. **Representative samples are better than biased samples.** As noted in Module 5, the best basis for generalizing is not from the exceptional and memorable cases one finds at the extremes but from a representative sample of cases. Research never randomly samples the whole human population. Thus, it pays to keep in mind what population a study has sampled.
2. **Less-variable observations are more reliable than those that are more variable.** As we noted in the example of the basketball player whose game-to-game points were consistent, an average is more reliable when it comes from scores with low variability.
3. **More cases are better than fewer.** An eager high school senior visits two university campuses, each for a day. At the first, the student randomly attends two classes and discovers both instructors to be witty and engaging. At the next campus, the two sampled instructors seem dull and uninspiring. Returning home, the student (discounting the small sample size of only two instructors at each institution) tells friends about the "great instructors" at the first school, and the "bores" at the second. Again, we know it but we ignore it: *Averages based on many cases are more reliable* (less variable) than averages based on only a few cases.

The point to remember: Smart thinkers are not overly impressed by a few anecdotes. Generalizations based on a few unrepresentative cases are unreliable.

When Is a Difference Significant?

Perhaps you've compared men's and women's scores on a laboratory test of aggression, and found a gender difference. But individuals differ. How likely is it that the gender difference you found was just a fluke? Statistical testing can estimate the probability of the result occurring by chance.

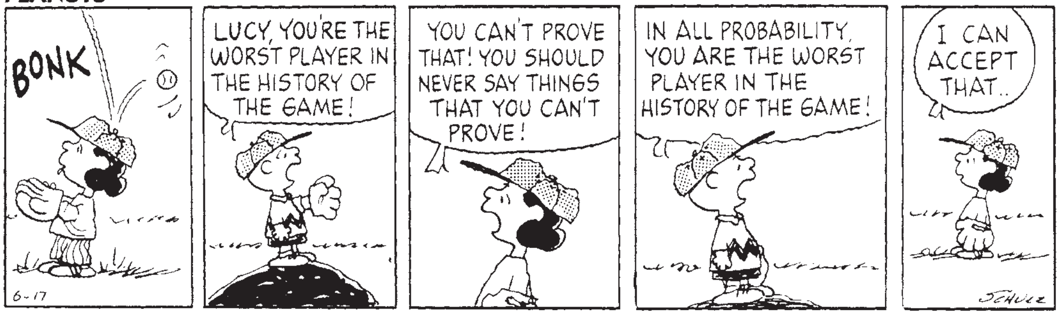
Here is the underlying logic: When averages from two samples are each reliable measures of their respective populations (as when each is based on many observations that have small variability), then their *difference* is likely to be reliable as well. (Example: The less the variability in women's and in men's aggression scores, the more confidence we would have that any observed gender difference is reliable.) And when the difference between the sample averages is *large,* we have even more confidence that the difference between them reflects a real difference in their populations.

In short, when sample averages are reliable, and when the difference between them is relatively large, we say the difference has **statistical significance.** This means that the observed difference is probably not due to chance variation between the samples.

statistical significance a statistical statement of how likely it is that an obtained result occurred by chance.

In judging statistical significance, psychologists are conservative. They are like juries who must presume innocence until guilt is proven. For most psychologists, proof beyond a

PEANUTS



Reprinted by permission of United Features Syndicate, Inc.

reasonable doubt means not making much of a finding unless the odds of its occurring by chance, if no real effect exists, are less than 5 percent.

When reading about research, you should remember that, given large enough samples, a difference between them may be "statistically significant" yet have little practical significance. For example, comparisons of intelligence test scores among hundreds of thousands of first-born and later-born individuals indicate a highly significant tendency for first-born individuals to have higher average scores than their later-born siblings (Kristensen & Bjerkedal, 2007; Zajonc & Markus, 1975). But because the scores differ by only one to three points, the difference has little practical importance.

The point to remember: Statistical significance indicates the *likelihood* that a result will happen by chance. But this does not say anything about the *importance* of the result.

AP® Exam Tip

Sometimes a phrase that is frequently used in the media has a more specific meaning when used in psychology. That's the case with the phrase "statistically significant." Make sure you know the precise meaning.

Before You Move On

▶ ASK YOURSELF

Find a graph in a popular magazine ad. How does the advertiser use (or abuse) statistics to make a point?

▶ TEST YOURSELF

Can you solve this puzzle?

The registrar's office at the University of Michigan has found that usually about 100 students in Arts and Sciences have perfect grades at the end of their first term at the University. However, only about 10 to 15 students graduate with perfect grades. What do you think is the most likely explanation for the fact that there are more perfect grades after one term than at graduation (Jepson et al., 1983)?

Answers to the Test Yourself questions can be found in Appendix E at the end of the book.

Module 7 Review

7-1 How do we describe data using three measures of central tendency, and what is the relative usefulness of the two measures of variation?

- A measure of central tendency is a single score that represents a whole set of scores. Three such measures are the *mode* (the most frequently occurring score), the *mean* (the arithmetic average), and the *median* (the middle score in a group of data).
- Measures of variation tell us how diverse data are. Two measures of variation are the *range* (which describes the gap between the highest and lowest scores) and the *standard deviation* (which states how much scores vary around the mean, or average, score).
- Scores often form a *normal* (or bell-shaped) *curve.*

7-2 How do we know whether an observed difference can be generalized to other populations?

- To feel confident about generalizing an observed difference to other populations, we would want to know that
  - the sample studied was representative of the larger population being studied;
  - the observations, on average, had low variability;
  - the sample consisted of more than a few cases; and
  - the observed difference was *statistically significant.*

Multiple-Choice Questions

1. Which of the following is a measure of variation?

a. Range
b. Mean
c. Mode
d. Frequency
e. Median

2. Which statistical measure of central tendency is most affected by extreme scores?

a. Mean
b. Median
c. Mode
d. Skew
e. Correlation

3. A researcher calculates statistical significance for her study and finds a 5 percent chance that results are due to chance. Which of the following is an accurate interpretation of this finding?

a. This is well beyond the range of statistical significance.
b. This is the minimum result typically considered statistically significant.
c. This is not statistically significant.
d. There is no way to determine statistical significance without replication of the study.
e. Chance or coincidence is unrelated to statistical significance.

4. Descriptive statistics __________, while inferential statistics __________.

a. indicate the significance of the data; summarize the data
b. describe data from experiments; describe data from surveys and case studies
c. are measures of central tendency; are measures of variance
d. determine if data can be generalized to other populations; summarize data
e. summarize data; determine if data can be generalized to other populations

5. In a normal distribution, what percentage of the scores in the distribution falls within one standard deviation on either side of the mean?

a. 34 percent
b. 40 percent
c. 50 percent
d. 68 percent
e. 95 percent

Practice FRQs

1. Explain the difference between descriptive and inferential statistics in research.

Answer (2 points)

1 point: Descriptive statistics organize and summarize the data collected during research.

1 point: Inferential statistics are used to help determine whether results can be generalized to a larger population through the calculation of statistical significance.

2. The following data set includes information from survey research in a psychology course regarding how many hours each individual in the class spent preparing for the exam.

Student	Amount of hours reported studying
1	2
2	3
3	6
4	8
5	9
6	9
7	21

Examine the data and respond to the following:

- What is the middle score in this distribution? What term is used to describe the middle score?
- What would be the most useful statistic for measuring the variation of the hours spent studying? Why is this statistic a better measure of variation than the range?

(3 points)

Module 8

Frequently Asked Questions About Psychology


wdstock/E+/Getty Images

Module Learning Objectives

8-1 Explain the value of simplified laboratory conditions in illuminating everyday life.

8-2 Discuss whether psychological research can be generalized across cultures and genders.

8-3 Explain why psychologists study animals, and describe the ethical guidelines that safeguard animal research participants.

8-4 Describe the ethical guidelines that safeguard human research participants.

8-5 Examine whether psychology is free of value judgments.

We have reflected on how a scientific approach can restrain biases. We have seen how case studies, naturalistic observations, and surveys help us describe behavior. We have also noted that correlational studies assess the association between two variables, which indicates how well one thing predicts another. We have examined the logic that underlies experiments, which use control conditions and random assignment of participants to isolate the effects of an independent variable on a dependent variable. And we have considered how statistical tools can help us see and interpret the world around us.

Yet, even knowing this much, you may still be approaching psychology with a mixture of curiosity and apprehension. So before we plunge in, let's entertain some frequently asked questions.

Psychology Applied

8-1 Can laboratory experiments illuminate everyday life?

When you see or hear about psychological research, do you ever wonder whether people's behavior in the lab will predict their behavior in real life? For example, does detecting the blink of a faint red light in a dark room have anything useful to say about flying a plane at night? If, after playing violent video games in the lab, teens become more willing to push buttons that they think electrically shock someone, does this indicate that playing shooter games makes someone more likely to commit violence in everyday life?

Before you answer, consider: The experimenter *intends* the laboratory environment to be a simplified reality—one that simulates and controls important features of everyday life. Just as a wind tunnel lets airplane designers re-create airflow forces under controlled conditions, a laboratory experiment lets psychologists re-create psychological forces under controlled conditions.

An experiment's purpose is not to re-create the exact behaviors of everyday life but to test *theoretical principles* (Mook, 1983). In aggression studies, deciding whether to push a button that delivers a shock may not be the same as slapping someone in the face, but the principle is the same. *It is the resulting principles—not the specific findings—that help explain everyday behaviors.*

When psychologists apply laboratory research on aggression to actual violence, they are applying theoretical principles of aggressive behavior, principles they have refined through many experiments. Similarly, it is the principles of the visual system, developed from experiments in artificial settings (such as looking at red lights in the dark), that researchers apply to more complex behaviors such as night flying. And many investigations show that principles derived in the laboratory do typically generalize to the everyday world (Anderson et al., 1999).

The point to remember: Psychological science focuses less on particular behaviors than on seeking general principles that help explain many behaviors. And remember: Although psychological principles may help predict behaviors for groups of people, they minimally predict behavior for any individual. Knowing students' grade level may clue us to their average vocabulary level, but individual students' word power will vary.

8-2 Does behavior depend on one's culture and gender?

What can psychological studies done in one time and place—often with people from what researchers call the WEIRD (*W*estern, *E*ducated, *I*ndustrialized, *R*ich, and *D*emocratic) cultures (Henrich et al., 2010) really tell us about people in general? As we will see time and again, **culture**—shared ideas and behaviors that one generation passes on to the next—matters. Our culture shapes our behavior. It influences our standards of promptness and frankness, our attitudes toward premarital sex and varying body shapes, our tendency to be casual or formal, our willingness to make eye contact, our conversational distance, and much, much more. *Collectivist* cultures, for example, emphasize group goals, while *individualist* cultures put a priority on individual goals. Being aware of such differences, we can restrain our assumptions that others will think and act as we do. Given the growing mixing and clashing of cultures, our need for such awareness is urgent.

culture the enduring behaviors, ideas, attitudes, values, and traditions shared by a group of people and transmitted from one generation to the next.

It is also true, however, that our shared biological heritage unites us as a universal human family. The same underlying processes guide people everywhere.

- People diagnosed with specific learning disorder (formerly called dyslexia) exhibit the same brain malfunction whether they are Italian, French, or British (Paulesu et al., 2001).
- Variation in languages may impede communication across cultures. Yet all languages share deep principles of grammar, and people from opposite hemispheres can communicate with a smile or a frown.
- People in different cultures vary in feelings of loneliness. But across cultures, loneliness is magnified by shyness, low self-esteem, and being unmarried (Jones et al., 1985; Rokach et al., 2002).


Alistair Berg/Alamy

Soccer shoes? Because culture shapes social behavior, actions that seem ordinary to some may seem odd to others. Yet underlying these differences are powerful similarities. Children everywhere love to play sports such as soccer. But many American children would only play with athletic shoes on a field, not barefoot in the street, as do these Burkina Faso boys.

We are each in certain respects like all others, like some others, and like no other. Studying people of all races and cultures helps us discern our similarities and our differences, our human kinship and our diversity.

You will see throughout this book that *gender* matters, too. Researchers report gender differences in what we dream, in how we express and detect emotions, and in our risk for alcohol use disorder, depression, and eating disorders. Gender differences fascinate us, and studying them is potentially beneficial. For example, many researchers believe that women carry on conversations more readily to build relationships, while men talk more to give information and advice (Tannen, 2001). Knowing this difference can help us prevent conflicts and misunderstandings in everyday relationships.

But again, psychologically as well as biologically, women and men are overwhelmingly similar. Whether female or male, we learn to walk at about the same age. We experience the same sensations of light and sound. We feel the same pangs of hunger, desire, and fear. We exhibit similar overall intelligence and well-being.

"All people are the same; only their habits differ." -Confucius, 551–479 B.C.E.

The point to remember: Even when specific attitudes and behaviors vary by gender or across cultures, as they often do, the underlying processes are much the same.

Ethics in Research

8-3 Why do psychologists study animals, and is it ethical to experiment on animals?

Many psychologists study animals because they find them fascinating. They want to understand how different species learn, think, and behave. Psychologists also study animals to learn about people. We humans are not *like* animals, we *are* animals, sharing a common biology. Animal experiments have therefore led to treatments for human diseases—insulin for diabetes, vaccines to prevent polio and rabies, transplants to replace defective organs.

"Rats are very similar to humans except that they are not stupid enough to purchase lottery tickets." -Dave Barry, July 2, 2002

Humans are complex. But the same processes by which we learn are present in rats, monkeys, and even sea slugs. The simplicity of the sea slug's nervous system is precisely what makes it so revealing of the neural mechanisms of learning. Sharing such similarities, should we not respect our animal relatives? "We cannot defend our scientific work with animals on the basis of the similarities between them and ourselves and then defend it morally on the basis of differences," noted Roger Ulrich (1991). The animal protection movement protests the use of animals in psychological, biological, and medical research. Researchers remind us that the animals used worldwide each year in research are but a fraction of 1 percent of the billions of animals killed annually for food. And yearly, for every dog or cat used in an experiment and cared for under humane regulations, 50 others are killed in humane animal shelters (Goodwin & Morrison, 1999).

Some animal protection organizations want to replace experiments on animals with naturalistic observation. Many animal researchers respond that this is not a question of good versus evil but of compassion for animals versus compassion for people. How many of us would have attacked Louis Pasteur's experiments with rabies, which caused some dogs to suffer but led to a vaccine that spared millions of people (and dogs) from agonizing death? And would we really wish to have deprived ourselves of the animal research that led to effective methods of training children with mental disorders, of understanding aging, and of relieving fears and depression? The answers to such questions vary by culture. In Gallup surveys in Canada and the United States, about 60 percent of adults deem medical testing on animals "morally acceptable." In Britain, only 37 percent do (Mason, 2003).

"Please do not forget those of us who suffer from incurable diseases or disabilities who hope for a cure through research that requires the use of animals." -Psychologist Dennis Feeney (1987)

Out of this heated debate, two issues emerge. The basic one is whether it is right to place the well-being of humans above that of animals. In experiments on stress and cancer, is it right that mice get tumors in the hope that people might not? Should some monkeys be exposed to an HIV-like virus in the search for an AIDS vaccine? Is our use and consumption of other animals as natural as the behavior of carnivorous hawks, cats, and whales? Defenders of research on animals argue that anyone who has eaten a hamburger, worn leather shoes, tolerated hunting and fishing, or supported the extermination of crop-destroying or plague-carrying pests has already agreed that, *yes,* it is sometimes permissible to sacrifice animals for the sake of human well-being.

Scott Plous (1993) notes, however, that our compassion for animals varies, as does our compassion for people—based on their perceived similarity to us. As Module 79 explains, we feel more attraction, give more help, and act less aggressively toward similar others. Likewise, we value animals according to their perceived kinship with us. Thus, primates and companion pets get top priority. (Western people raise or trap mink and foxes for their fur, but not dogs or cats.) Other mammals occupy the second rung on the privilege ladder, followed by birds, fish, and reptiles on the third rung, with insects at the bottom. In deciding which animals have rights, we each draw our own cut-off line somewhere across the animal kingdom.

If we give human life first priority, what safeguards should protect the well-being of animals in research? One survey of animal researchers gave an answer. Some 98 percent supported government regulations protecting primates, dogs, and cats, and 74 percent supported regulations providing for the humane care of rats and mice (Plous & Herzog, 2000). Many professional associations and funding agencies already have such guidelines. British Psychological Society guidelines call for housing animals under reasonably natural living conditions, with companions for social animals (Lea, 2000). American Psychological Association (APA) guidelines state that researchers must ensure the "comfort, health, and humane treatment" of animals and minimize "infection, illness, and pain" (APA, 2002). The European Parliament now mandates standards for animal care and housing (Vogel, 2010).

"The greatness of a nation can be judged by the way its animals are treated." -Mahatma Gandhi, 1869–1948

Animals have themselves benefited from animal research. One Ohio team of research psychologists measured stress hormone levels in samples of millions of dogs brought each year to animal shelters. They devised handling and stroking methods to reduce stress and ease the dogs' transition to adoptive homes (Tuber et al., 1999). Other studies have helped improve care and management in animals' natural habitats. By revealing our behavioral kinship with animals and the remarkable intelligence of chimpanzees, gorillas, and other animals, experiments have also led to increased empathy and protection for them. At its best, a psychology concerned for humans and sensitive to animals serves the welfare of both.

AP Photo/Mary Altaffer

Animal research benefiting animals Thanks partly to research on the benefits of novelty, control, and stimulation, these gorillas are enjoying an improved quality of life in New York's Bronx Zoo.

8-4 What ethical guidelines safeguard human participants?

Does the image of white-coated scientists delivering electric shocks trouble you? If so, you'll be relieved to know that most psychological studies are free of such stress. With people, blinking lights, flashing words, and pleasant social interactions are more common. Moreover, psychology's experiments are mild compared with the stress and humiliation often inflicted by reality TV shows. In one episode of *The Bachelor,* a man dumped his new fiancée—on camera, at the producers' request—for the woman who earlier had finished second (Collins, 2009).

Occasionally, though, researchers do temporarily stress or deceive people, but only when they believe it is essential to a justifiable end, such as understanding and controlling violent behavior or studying mood swings. Some experiments won't work if participants know everything beforehand. (Wanting to be helpful, the participants might try to confirm the researcher's predictions.)

Ethical principles developed by the American Psychological Association (2010), by the British Psychological Society (2009), and by psychologists internationally (Pettifor, 2004), urge researchers to (1) obtain potential participants' **informed consent,** (2) protect them from physical or emotional harm and discomfort, (3) keep information about individual participants confidential, and (4) fully **debrief** people (explain the research afterward). Moreover, most universities (where a great deal of research is conducted) now have an ethics committee—an Institutional Review Board (IRB)—that screens research proposals and safeguards participants' well-being.

The ideal is for a researcher to be sufficiently informative *and* considerate so that participants will leave feeling at least as good about themselves as when they came in. Better yet, they should be repaid by having learned something.

informed consent an ethical principle that research participants be told enough to enable them to choose whether they wish to participate.

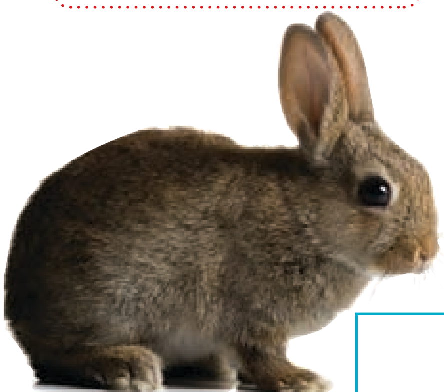
debriefing the postexperimental explanation of a study, including its purpose and any deceptions, to its participants.

8-5 Is psychology free of value judgments?

Psychology is definitely not value-free. Values affect what we study, how we study it, and how we interpret results. Researchers' values influence their choice of topics. Should we study worker productivity or worker morale? Sex discrimination or gender differences? Conformity or independence? Values can also color "the facts." As we noted earlier, our preconceptions can bias our observations and interpretations; sometimes we see what we want or expect to see (**FIGURE 8.1**).

Even the words we use to describe something can reflect our values. In psychology and in everyday speech, labels describe and labels evaluate: One person's *rigidity* is another's *consistency*. One person's *faith* is another's *fanaticism*. One country's *enhanced interrogation techniques,* such as cold-water immersion, become *torture* when practiced by its enemies. Our labeling someone as *firm* or *stubborn, careful* or *picky, discreet* or *secretive* reveals our own attitudes.

"It is doubtless impossible to approach any human problem with a mind free from bias."
-Simone de Beauvoir, *The Second Sex*, 1953



Mike Kemp/Getty Images

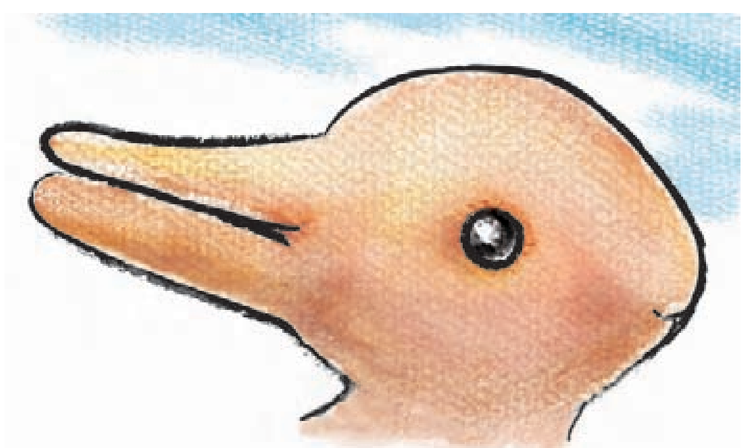


Figure 8.1
What do you see? Our expectations influence what we perceive. Did you see a duck or a rabbit? Show some friends this image with the rabbit photo above covered up and see if they are more likely to perceive a duck head instead. (From Shepard, 1990.)

Popular applications of psychology also contain hidden values. If you defer to "professional" guidance about how to live—how to raise children, how to achieve self-fulfillment, what to do with sexual feelings, how to get ahead at work—you are accepting value-laden advice. A science of behavior and mental processes can help us reach our goals. But it cannot decide what those goals should be.

If some people see psychology as merely common sense, others have a different concern—that it is becoming dangerously powerful. Is it an accident that astronomy is the oldest science and psychology the youngest? To some, exploring the external universe seems far safer than exploring our own inner universe. Might psychology, they ask, be used to manipulate people?

Knowledge, like all power, can be used for good or evil. Nuclear power has been used to light up cities—and to demolish them. Persuasive power has been used to educate people—and to deceive them. Although psychology does indeed have the power to deceive, its purpose is to enlighten. Every day, psychologists are exploring ways to enhance learning, creativity, and compassion. Psychology speaks to many of our world's great problems—war, overpopulation, prejudice, family crises, crime—all of which involve attitudes and behaviors. Psychology also speaks to our deepest longings—for nourishment, for love, for happiness. Psychology cannot address all of life's great questions, but it speaks to some mighty important ones.


Library of Congress

Psychology speaks In making its historic 1954 school desegregation decision, the U.S. Supreme Court cited the expert testimony and research of psychologists Kenneth Clark and Mamie Phipps Clark (1947). The Clarks reported that, when given a choice between Black and White dolls, most African-American children chose the White doll, which seemingly indicated internalized anti-Black prejudice.

Before You Move On

▶ ASK YOURSELF

Were any of this module's Frequently Asked Questions your questions? Do you have other questions or concerns about psychology?

▶ TEST YOURSELF

How are human and animal research participants protected?

Answers to the Test Yourself questions can be found in Appendix E at the end of the book.

Module 8 Review

8-1 Can laboratory experiments illuminate everyday life?

- Researchers intentionally create a controlled, artificial environment in the laboratory in order to test general theoretical principles. These general principles help explain everyday behaviors.

8-2 Does behavior depend on one's culture and gender?

- Attitudes and behaviors may vary somewhat by gender or across *cultures*, but because of our shared human kinship, the underlying processes and principles are more similar than different.

8-3 Why do psychologists study animals, and is it ethical to experiment on animals?

- Some psychologists are primarily interested in animal behavior; others want to better understand the physiological and psychological processes shared by humans and other species.
- Government agencies have established standards for animal care and housing. Professional associations and funding agencies also establish guidelines for protecting animals' well-being.