

# 4

## CRITICAL THINKING AND SCIENTIFIC REASONING



### LEARNING OUTCOMES

After studying this chapter, you should be able to:

1. Explain how nonscientific and scientific reasoning differ.
2. Identify the strengths and weaknesses of various methods used as evidence in scientific research and argumentation.
3. Determine whether a cause-and-effect relationship is likely to exist.
4. Recognize the limits of the scientific approach and inductive reasoning.
5. Recognize thinking errors associated with inductive and scientific reasoning.

### CHAPTER OUTLINE

What Do You Think?

Comparing Scientific and Nonscientific Approaches

Science as an Approach to Knowledge and Evidence

Causation and the Quality of Scientific Evidence

Strengths and Weaknesses of Research Methods

Thinking Errors in Wrongly Inferring Causation

True Experiments and Causation

What Is Scientific Thinking?

Summary

Review Questions

### WHAT DO YOU THINK?



AP Photo

**FIGURE 4.1** Poster for Heaven's Gate, a UFO cult, announcing the end of their time on Earth, detailing when group members would meet with aliens from outer space who resided in a different dimension.

In 1997, 39 members of the Heaven's Gate cult committed suicide by drinking vodka laced with cyanide so they could rendezvous with alien beings whose spaceship was approaching Earth behind Comet Hale-Bopp. Figure 4.1 shows one of the cult's posters announcing this event. Cult members believed that if they committed suicide, alien beings from another dimension would open the gates to heaven for them, thereby saving them

from Earth's impending destruction. Some members had belonged to the cult for decades, whereas others were fairly new to these strange beliefs. Yet they all decided to commit suicide. Why?



### Practice Thinking 4.1: What Do You Think?

Please explain how you know.

1. When people join a cult, do they accept all the cult's beliefs and lose their ability to think critically? How could you find out?

---

2. Does observing aggressive behaviors in movies, TV, or video games cause a person to behave aggressively? How could you find out?

---

3. Does having been sexually abused as a child lead to personality problems as an adult and cause the victim to later become a sexual abuser?

---

4. Can science answer all questions? Why or why not?

---

In this chapter, we examine how science helps us obtain better answers to these fascinating and important questions than do personal experience, anecdotes, and other nonscientific approaches. As you will see, the methods that scientists use are superior.

### COMPARING SCIENTIFIC AND NONSCIENTIFIC APPROACHES

Scientific evidence, unlike the informal and unsystematic observations of nonscientific evidence, is based on careful and systematic observation. For example, scientists who want to conduct rigorous studies of precognition would use carefully controlled experimentation, not informal observation such as that

used by some people at Duke University, who mistakenly concluded that Lee Fried could predict the future. This reliance on personal experience provided no systematic way to test the quality of the information obtained. In contrast, because the scientific approach employs rules and strategies for reasoning effectively about carefully made observations, scientists are able to evaluate and even improve the quality of their data.

Science is called an *empirical approach* because of its reliance on verifiable observations, or data that can be shown to have a certain quality. Scientists use **scientific standards of evidence**, which are rules and principles used to determine higher- versus lower-quality scientific evidence. The sciences show overlap in the standards they use, but each discipline may develop its own specific rules of reasoning (Bensley, 2011). For instance, all the sciences agree on the general rule that theories must be consistent with evidence. Psychologists specify that experiments can better demonstrate cause and effect than other research designs can. They develop even more specific rules for interpreting the data from research instruments and techniques, such as brain-scanning equipment, that provide much more precise data than nonscientific evidence does. Proper use of the rules for reasoning well about carefully made observations gives the scientific approach its power to answer many different questions.

### SCIENCE AS AN APPROACH TO KNOWLEDGE AND EVIDENCE

To answer the many questions posed in a particular field, the scientist looks for lawful relations among variables by using specific methods and techniques based on rules for reasoning effectively about data. For example, suppose a social psychologist wants to find out why a person is helpful in one situation but not in another. She would study the variable—helpfulness—under various conditions. A **variable** is a characteristic or event of interest that can take on different values. In everyday, nonscientific terms, you might imprecisely refer to someone as “selfish” or “helpful.” The psychologist would more precisely define the variable of helpfulness, in terms of how many times the research participant was observed to help. Or the researcher might rate the participant on a selfishness scale ranging from 1 = *not at all helpful* to 7 = *extremely helpful*. This illustrates two ways to **operationalize**, or represent, the variable in terms of methods, procedures, and measurements. Assigning a number to a helpfulness scale is a useful first step in describing a single variable, but it is limited in what it tells us. To validly measure the construct underlying helpfulness, scientists would use other reliable gauges that could provide converging information about it (Grace, 2001).

Understanding something complex, such as helping behavior, requires that we understand the relationships between variables, not just one variable. A **relationship** between variables indicates that the values of one variable change consistently in relation to the values of another variable. Scientists seek to express relationships in precise terms that can be observed. For instance, on average, participants who score low—say, a 2—on the helpfulness scale tend to score higher on the amount of time spent talking about themselves in a 10-minute conversation—say, 8 out of 10 minutes.

Scientists refer to the predicted relationship between two or more variables as a **hypothesis**. More formally, a hypothesis is often deduced from a theory in the form of a specific prediction, as discussed in Chapter 2. It is a claim about what will happen if we assume that some theory is true. For example, from the general theory that people are basically selfish and motivated by self-interest, we might predict that if participants have the opportunity to help someone else in a new situation, they will not help. Hypotheses can also originate from other sources, including personal experience.

A hypothesis typically makes a prediction about one of two types of relationships: (1) an association or (2) a cause-and-effect relation. In an association, sometimes referred to as a *correlation*, the values of two variables are simply related or change together in a consistent way. In a positive association, as one variable increases, the other variable increases along with it; or as one variable decreases, the other variable tends to decrease. For instance, the more helpful people are, the more likely they are to volunteer. Putting this hypothesis in terms of scores on a 7-point helpfulness rating scale, we might say, “We predict that the higher a person’s helpfulness rating score, the more often that person will volunteer to help.” It is also true in this positive association that the lower the score on the helpfulness scale, the lower the tendency to volunteer—which illustrates that in a positive association, the values of two variables change in the *same* direction.

In a negative association between variables, the values of the two variables consistently change together in the *opposite* direction. Stating this as a correlational hypothesis for a research study, we might say, “We expect that the higher a participant’s score on the 7-point helpfulness scale, the less time that participant will spend doing something to benefit himself or herself.”

Showing a cause-and-effect relationship takes more than simply showing an association between variables. To show causation, changes in one variable (the cause) must occur *before* changes occur in the other variable (the effect). Consider this causal hypothesis: “If one group studies a list of words for longer than another group does, then the group that studied longer will recall more on a

test of the new words.” In this hypothesis, how long people study (the cause) must happen before the effect (how many new words they learned). It makes no sense for a causal form of this hypothesis to predict that people will first recall the new list of 30 words and then will study the words for a longer or shorter time. They are not new words if one has already recalled them.

People find relationships between variables in their daily lives, not just in scientific research. But how good are people at assessing whether variables are correlated? Look at the data in Table 4.1, showing the frequency of cases in which an abnormal behavior is either present or not present in relation to the moon’s phase (full or not full). Suppose, as shown in cell A, that people working at a mental health facility observed 12 cases in which the moon was full and people behaved abnormally. Do the data in the fourfold table show that the presence of a full moon is related to abnormal behavior?

**TABLE 4.1 The Fourfold Table Showing Cell Frequencies**

	Full Moon	No Full Moon
Abnormal behavior	12 (Cell A)	6 (Cell B)
No abnormal behavior	6 (Cell C)	3 (Cell D)

Examining the data in Table 4.1, many of the mental health facility staff mentioned in Chapter 1 would likely find a correlation between the full moon and abnormal behavior, concluding that people tend to behave abnormally during a full moon. But they would be mistaken. This thinking error occurs because people tend to notice co-occurrences of events, like those shown in the higher frequencies of cell A, and do not take into account the frequencies in other cells of the table. You are not likely to hear people say, “Hey, there’s a full moon tonight, but nobody is behaving strangely!” (Kohn, 1990). This demonstrates how people often fail to take into account the six cases in cell C who are *not* behaving abnormally during a full moon. Nor are people likely to say, “Wow, the patients are behaving abnormally, but there’s not even a full moon!” which demonstrates how they tend to ignore the six cases in cell B. We must take these cases into account because they provide evidence that no relation exists. What does the research say about people’s ability to analyze data like these?

The research shows that people are not very good at assessing these kinds of associations; they pay too much attention to cell A and neglect the other cells (Smedsland, 1978). This thinking error is called **illusory correlation** because

people mistakenly perceive a relationship (correlation) between two variables when none exists. For example, people often find illusory correlations between certain traits and specific groups, leading to the formation of stereotypes that are unfairly applied to individuals. Even professionals sometimes fall prey to illusory correlation, as when clinicians incorrectly diagnose mental disorders by paying attention to co-occurrences of expected features (cell A) and not considering all the information available (Chapman & Chapman, 1967).

Taking a scientific approach to forming and testing hypotheses can help safeguard against illusory correlation and other thinking errors. By systematically making observations that could show whether variables are changing together or not, scientists look at the data in *all* four cells of a fourfold table. In this regard, a good scientific hypothesis is **falsifiable**, or can be shown to be false (Popper, 1959). When scientists think critically, they do not seek to confirm a hypothesis. Instead, they set up their test to also make observations that could *disconfirm* the hypothesis. Then they examine *all* the data—those that support the hypothesis and those that do not. If we were unable to show that a hypothesis was false, how could we ever find out that our hypotheses and theories were wrong and needed correction? In other words, falsifiability allows for **self-correction** in science (Myers & Hansen, 2012).

For example, the hypothesis “If people study this critical thinking book enough, they will be able to think critically in a new situation” is not falsifiable. Using the qualifier “enough” is not specific and would always allow for other after-the-fact explanations that would prevent it from being disconfirmed. If people studied this book but were not able to think critically in a new situation, we could always say, “They did not study it enough.” But what is *enough*? A good scientific hypothesis needs to make a specific prediction.

When scientists conduct a study that supports a theory or hypothesis, they often try to **replicate** the study, attempting to repeat the observations under similar conditions. A positive outcome can provide more inductive support for a theory, but replicating the findings of a previous study never proves the theory is true. If a theory or hypothesis survives even more rigorous testing under conditions that could disconfirm it, then these positive results can strengthen it even more. On the other hand, if a good experiment fails to support social learning theory and this negative finding is replicated with other high-quality studies, eventually we would decide that under some specific condition, the theory is false and should be revised or is even completely wrong and should be rejected. This outcome might disturb some, but it is a fundamental way for us to correct mistaken scientific ideas.

After our discussion of the neglect of negative evidence in illusory correlation, it may not surprise you that people often do not seek evidence that could falsify their own theories when testing them. Once people find a relationship or form a belief, they tend to look for evidence that confirms or supports their favored belief, often ignoring or minimizing evidence that could disconfirm it; this illustrates a thinking error called **confirmation bias**. In one study, Lord, Ross, and Lepper (1979) first asked participants if they favored or opposed the death penalty. Participants were then given the results of two experiments of equal quality, one that supported the idea that the death penalty deterred further crime and another that did not support this claim. Consistent with the effect of confirmation bias, participants rated the results of the study supporting what they already believed as more convincing than the results of the study that disagreed with their position. After they read the results of the study that disagreed with their position, this negative evidence should have made participants less convinced that they were right, but surprisingly, they became even more convinced of their position.

Confirmation bias is a very common thinking error that affects many kinds of judgments (Nickerson, 1998) and that we will examine at various points in this book. For instance, one study of juror reasoning showed that as a trial progressed, prospective jurors tended to bias their interpretation of newly presented evidence to be consistent with whatever their current preference was in terms of the defendant's guilt or innocence (Carlson & Russo, 2001). In another study, party affiliation biased ballot counters to judge ambiguous ballots as votes in favor of candidates in their party (Kopko, Bryner, Budziak, Devine, & Nawarra, 2011). An everyday example is the "toupee fallacy," which shows confirmation bias in people who believe they can detect when a man is wearing a toupee. Whenever these people learn that someone they suspect of wearing a toupee is actually wearing one, this confirms their belief in their ability. But, because they do not systematically check whether other men they think are *not* wearing toupees are actually wearing them, they get a biased estimate of their detection ability (Novella, 2012).

Are scientists immune to confirmation bias? No, but they have strategies for countering it, such as the **peer review of research**. When a scientist submits his or her research to be accepted for publication, an editor sends the manuscript to experts on the question (the scientist's peers) who evaluate the quality of the research and look for problems and ways the researcher's conclusions could be false. The peer review process is not foolproof, however, and can itself be subject to confirmation bias. Mahoney (1977) asked scientific journal reviewers to evaluate manuscript submissions of studies that were identical except for their results. The reviewers gave higher ratings to the manuscripts with results supporting

their own favored theories than to manuscripts with results that challenged their favored views. This, of course, is problematic. Fortunately, peer review still works because others in the scientific community may find fault with a study or will be unable to replicate the findings of the study in question.

Probability is another tool scientists use to decide whether relationships are real or illusory. **Probability** is “the likelihood that a particular event or relation will occur” (Vogt, 1993, p. 178). For instance, to determine how likely a group rated as selfish and a group rated as unselfish will engage in helping behavior, a researcher conducts statistical analyses on some measure of helping behavior. This involves using probability to estimate the likelihood of obtaining some difference between the two groups simply by chance. If the researcher finds a very low probability that the observed difference between the selfish and unselfish groups was due simply to chance (e.g., if the difference would occur by chance fewer than 5 times out of 100), then the researcher concludes it is likely that a real difference in the two groups exists. The researcher declares this difference to be **statistically significant**, or just significant. This significant difference suggests that a real relationship exists between ratings of people’s selfishness and their willingness to help, supporting the hypothesis that unselfish people help more. This also helps reduce the probability that any observed difference in the groups was just a random one that might have been simply observed by chance (as in an illusory correlation).

### CAUSATION AND THE QUALITY OF SCIENTIFIC EVIDENCE

Different scientific research methods and designs provide evidence that varies in quality. Because the methods of science involve using observation to test hypotheses and to evaluate theories, the quality of the evidence offered by scientific research depends on the ability to collect high-quality data. The best evidence comes from studies in which observations were made with objectivity, without error, and under carefully controlled conditions.

The quality of the evidence provided by scientific research methods also depends on the degree to which a particular method can establish a causal relation between variables. Recall that the goals of psychology as a science are to describe, predict, explain, and control or manipulate behavior. In order to reach the important goal of explaining behavior, we must be able to show the causes of behavior. When we speak of a **cause**, we are referring to something that has produced an effect. A cause precedes the event it produces (the effect). Knowing the

cause can help explain why the effect happened (Zechmeister & Johnson, 1992). To better understand how something might be the cause of some behavior, let's look closely at the three **criteria for establishing causation**, shown in Table 4.2. Recall that a criterion is a standard that must be met or a condition that must be present in order to confirm that something is true.

**TABLE 4.2 Three Criteria for Establishing Causation**

1. Two events must covary or vary together consistently (covariation).
2. One event must occur before the other (time order).
3. Plausible alternative explanations for the covariation must be eliminated.

To illustrate the use of these criteria, let's apply them to the question of whether precognition caused the supposedly correct prediction of the plane crash in the Lee Fried example from Chapter 3. To show that the covariation criterion was met, we would have to demonstrate that Fried's precognitive ability and the predicted event changed together. We would have to prove that when Fried had his premonition, it was systematically related to the crash. The two events *appear* to have occurred close together in time, suggesting that covariation was present. Also, it *appears* that the precognition occurred before the letter was delivered, suggesting that the criterion of time order had been met, although no other verification of this criterion was demonstrated. Finally, Fried's handing over a sealed letter to a public figure suggested that the letter would not be tampered with and that the event could be documented. This *appears* to eliminate the alternative explanation of cheating as the cause of the correct prediction.

A closer examination of the events shows that none of the criteria were actually met. Fried's confession that he was a magician, which implies that he used deception, suggests that the two events that actually covaried were (1) the swapping of the letters and (2) the reading of the new letter, which was presumed to be the original letter with the prediction. But the contents of the letter had been put in the envelope *after* the jumbo jet crashed, so no premonition occurred before the event. Thus, time order was not met. Nor were two plausible alternative explanations eliminated. First, someone should have tried to falsify Fried's claim by checking how many other events he had predicted accurately. Psychics frequently guess, so sometimes their predictions do turn out to be right, simply by chance. The second, much more plausible, explanation someone should've checked before concluding that precognition was the cause is that Fried engaged

in trickery. For example, did Fried ever have access to the letter after the crash occurred? Only Fried's word supported his use of precognition, and he later cast doubt on his claim that he had precognitive ability.

It is clear from analysis of this example that it is virtually impossible to demonstrate causation in an anecdote. In fact, only the true experiment, one of the scientific research designs we discuss next, can put us in the position to infer causation.

### STRENGTHS AND WEAKNESSES OF RESEARCH METHODS

Various research methods differ in the kind and quality of information they can provide. In particular, they differ in the extent to which they can control extraneous variables and the degree to which causal inferences can be made. Table 4.3 (on page 90) shows the strengths and weaknesses of various commonly used research methods. One important idea in Table 4.3 is that only the true experiment allows the researcher to make causal inference because it's the only method through which all the criteria for causation can be met; the research methods described next do not have that same capability.

A **case study** provides a detailed description of part of an individual's life, often documenting a person's abilities, traits, symptoms, behaviors, and treatment. Case studies are often used to study the behavior of people in treatment, as well as special individuals with certain traits and abilities. Although covariation of variables can sometimes be shown, it is much more difficult to show that one variable precedes another. It is frequently even more difficult to eliminate other variables that could be having an effect besides the one that is supposed to be the cause. We call these **extraneous variables** because they are "extra" variables outside the intended focus of our study that could provide alternate explanations for the research findings.

Compared with anecdotes that also often involve a description of a single person in a situation, case studies differ in that the observations are made systematically and are based on recordings of observations, not simply someone's recollection. Although multiple observations can be made of a single person, like anecdotes, case studies still tend to suffer in terms of quantity of evidence.

**Survey research** involves asking participants multiple questions. Questions can be asked in a mailed or emailed questionnaire or in a face-to-face interview. The main advantage of surveys and questionnaires is that many questions can be asked of many people. Questionnaires are thus economical and versatile and can address a wide variety of topics, traits, and experiences.

But surveys and questionnaires can present difficulties too. Gathering reliable data depends on the wording of the questions, how the participants are selected, and the truthfulness of their responses. In addition, if researchers do not select respondents in a way that makes the sample representative of a population or if many participants do not respond, it will be difficult to generalize validly from those in the sample responding to the population at large. Also, respondents may not be honest in reporting their opinions or may not remember factual information accurately when asked. Although covariation may be shown between one item and another, the other two criteria for causation are typically not met. In particular, it is very difficult to control extraneous variables—such as individual differences in respondents—that could affect how respondents answer.

Further limitations of surveys are that responses to survey questions can be sensitive to placement in the survey, the context in which questions are asked, and the wording and form of a question (Schwartz, 2007). For example, when Schuman and Presser (1981) asked respondents to report what they considered “the most important thing for children to prepare them for life,” only 4.6% wrote a response on the order of “think for themselves.” Yet when “think for themselves” was put in a list of alternatives, 61.5% of the participants selected it. This finding suggests that asking a question in two different ways can yield wildly different answers. Ironically, it also suggests that when people were asked to think for themselves by generating a response, they tended *not* to answer that “thinking for themselves” was important. However, when they didn’t have to think of the response on their own, they more often thought that “thinking for themselves” was important.

Using **field studies** and **naturalistic observation**, researchers collect data in the natural environment, which has the advantage of avoiding the artificiality of laboratory research. Jane Goodall and other primatologists have used naturalistic observation to study chimpanzees in the wild. Developmental psychologists have used it to study the interactions of children in their classrooms using video cameras to record their behaviors for later study. Because observed behaviors are simply part of a stream of behavior, it is difficult to establish the time order of behaviors and nearly impossible to eliminate extraneous variables. Suppose you are observing the selfish and unselfish behaviors of a person in everyday situations. The person may respond to a variety of situational variables, such as rewards for being generous, that also affect the apparently unselfish behavior.

A **participant–observer** study is a type of field study, often used in social psychology and sociology, in which a researcher infiltrates a group to study it without the participants knowing they are being studied. This reduces the reactivity of

participants or their tendency to behave a certain way because they know they are being observed. The sociologist Robert Balch became a participant–observer, joining the Heaven’s Gate cult to study how new members were recruited and came to accept the group’s unusual beliefs (Balch, 1993). This method allowed Balch and his graduate student assistant David Taylor to find out whether new members accepted all of the cult’s extraordinary claims when they first joined, thereby losing their ability to think critically about the claims.

Balch and Taylor found that new members were often seekers who focused on a spiritual quest for fulfillment and had been part of other groups with paranormal beliefs. Consequently, conversion to the cult was not dramatic but seemed like the next logical step for some. Other evidence that members had not been brainwashed came from the fact that many converts left the cult soon after joining and that even some longtime members sometimes raised doubts about the group’s fundamental beliefs (Balch, 1993; Balch & Taylor, 1977). Although naturalistic observation by participant–observers made these interesting and important findings possible, this method would not allow the researchers to determine the *causes* of cult-member behavior.

**Correlational studies** seek to find a *quantitative* relationship between two or more variables in which the variables covary, or vary together. As mentioned earlier, positive correlations occur when two variables vary together in the same direction—as when empathy and the willingness to help are positively correlated. A person feeling empathy identifies with and feels what another person is feeling and tends to help more. A negative correlation might be found between empathy and selfishness. As selfishness increases, people tend to feel less empathy. These variables are covarying in the opposite direction.

Obviously, covariation is easy to establish if the two variables are correlated; however, time order is more difficult to show. Although it appears that an individual may first empathize with someone in need of help and then help that person, the relation between the two variables might go in the other direction. It may be that helping another person reinforces feelings of empathy, making helpers feel more connected to the people they have helped.

Correlational studies cannot easily eliminate other plausible alternative explanations, either, because another variable may actually be the cause for change in the two correlated variables; this is sometimes called the *third-variable problem*. Suppose we find a significant correlation between watching violence on TV and subsequent aggression. We might be tempted to conclude that watching violence on TV leads to more aggressive behavior, but what if a third variable, such as “liking to observe violence,” leads people to be both more aggressive and

to watch more violence on TV? It is difficult, if not impossible, to eliminate this third variable as an explanation in a correlational study. Consequently, correlation does not allow us to infer cause and effect.

### THINKING ERRORS IN WRONGLY INFERRING CAUSATION

You may have heard the expression, “Correlation does not imply causation.” Inferring that one of two simply correlated variables is a cause of the other is a thinking error called **confusing correlation with causation**. A good example, discussed in Chapter 2, is the misconception in the 1980s that improving self-esteem would improve academic performance. The many attempts to improve students’ academic performance by raising their self-esteem were largely unsuccessful. Although self-esteem is modestly correlated with academic achievement, it does not *cause* it (Baumeister, Campbell, Krueger, & Vohs, 2003). Those who perform better in school are simply more likely to feel better about themselves.

Another misconception related to confusing correlation with causation is the popular belief that victims of sexual abuse will necessarily develop personality problems in adulthood and will become abusers themselves. Although sexual abuse is indeed too common and can be very harmful, the research generally does not support that it *causes* people to develop a specific set of personality issues, such as low self-confidence and problems with intimacy and relationships, that victims carry for the rest of their lives (Lilienfeld, Lynn, Ruscio, & Beyerstein, 2010). Rather, the research shows that abused people are generally resilient and able to adjust to the early trauma. In a meta-analysis of many studies of college students, Rind, Bauserman, and Tromovitch (1998) found that although the students’ experience of sexual abuse was related to some psychological problems later in life, the correlations were low.

Moreover, research by Salter and colleagues (2003) found that less than 12% of men who had been sexually abused as children later became abusers themselves (Salter et al., 2003). When compared with the approximately 5% of men who were *not* abused but who later committed sexual abuse, the frequency for sexual abuse victims is certainly higher; but it also means that about 88% of sexual abuse victims do *not* become abusers themselves. Salter and colleagues also found that other risk factors were often present in the abusers as children, such as a lack of supervision and having witnessed serious violence among family members, which may have caused the later abusive behavior. This further suggests that although a correlation is present, causation should not be inferred.

Another kind of thinking error about causation, called **post hoc reasoning**, occurs when people incorrectly assume that something that merely happened to occur before an event was the actual cause of the event. The English translation of *post hoc* is “after this”; it comes from the longer Latin expression *post hoc, ergo propter hoc*, which means “after this, therefore because of this.” Both expressions refer to making an unwarranted assumption about time order, specifically that something occurring *after this* other event was caused by it.

To illustrate, suppose you begin taking vitamins and later notice that your concentration is better than before you started taking them. From this, you may mistakenly conclude that the vitamins caused the improvement in your concentration. Looking back at the events, it may seem that taking the vitamin was the first event and led to better concentration, but the two events may be simply coincidental, and the criterion of time order has not been established. Nor have you met the other criterion of eliminating plausible, alternative explanations. You may have expected the vitamins to improve your well-being in general (a placebo effect, discussed in Chapter 5), and you likely did not establish a controlled variable against which to measure any effects. Or what if, during this time, you also exercised more or got more sleep? These could be the actual causes of the perceived change in your concentration. What is needed is a method that allows us to manipulate one variable so that it clearly occurs before the other variable while we also control other potential causes.

### TRUE EXPERIMENTS AND CAUSATION

**True experiments** do allow us to make causal inferences because all three criteria of causation can be met. In a true experiment, an **independent variable** is manipulated—a variable that the researcher wants to demonstrate is the cause of an effect in another variable called the dependent variable. The **dependent variable** is the measured variable. In psychological research, the dependent variable is almost always some behavior.

For instance, an experimenter could test the commonsense idea that there is strength in numbers. You might expect that if you needed help, you would be more likely to receive it when more people are available to help than when just one person knows of your plight. An experimenter could test this hypothesis by varying the levels of the independent variable in terms of an expectation of how

many people the research participants believed were available to help a person in need. He could randomly assign participants to two groups: One group is led to believe that only each of them individually is available to help; a second group is led to believe that others are also available to help.

After manipulating the independent variable, the experimenter then measures the effect on the dependent variable—the willingness to help the person in need. Because these two levels were presented before the dependent variable, the criterion of time order has been met. If the independent variable produces significant differences in the two groups' willingness to help, then the criterion of covariation has also been met. Finally, because of random assignment of participants to the two groups, the experimenter has controlled extraneous variables, such as individual differences in the tendency to help, and has thus met the criterion of elimination of plausible alternative explanations.

Note that true experiments are useful not only for showing causation but also for correcting misconceptions. When Darley and Latané (1968) conducted this experiment on helpfulness, they found that participants who thought they alone were available to help were more likely to provide that help. This revealed the real cause of helping in this situation, contradicting the misconception that someone in need is more likely to receive help when more people are present (as discussed further in Chapter 10).

The goal of the experimenter is to show that it was the independent variable—and no other variable—that had an effect on the dependent variable. Experimental controls help to accomplish this goal, at which point the experiment is said to have **internal validity**. When internal validity is high, the experimenter is in a good position to show that it was the independent variable and no other variable that caused the changes in the dependent variable. This result is possible because this true experiment was well conducted and controlled for extraneous variables that could have caused the changes in the dependent variable.

However, what if an experimenter tests participants in a way that allows an extraneous variable to bias the results? Returning to our helpfulness experiment, suppose the researcher always tested the group that thought no one else was available to help before testing the group that thought others were available to help? Suppose further that he always tested first the participants who had signed up for the experiment first and thus seemed most eager to assist? This would introduce a confounding variable, willingness to help, an alternate explanation for why participants helped (i.e., other than the independent variable

being the reason). How would the experimenter know whether the hypothesis had been supported?

Even if the group members who thought they alone knew the plight of the person were found to help more than the other group, we could not be sure that it was because they thought they alone were available to help. It could be that the group tested first was already more likely to assist because they volunteered first, suggesting greater willingness to help. We say that whether or not anyone else was aware of the plight of the person (the independent variable) was confounded with “time of volunteering/willingness to help.” A **confounding** variable is an extraneous variable that varies along with the independent variable and could also plausibly account for the changes in the dependent variable. This threatens the experiment’s internal validity because other variables, such as the helpfulness of volunteers (rather than availability of others to help), could be the real cause of the changes in the dependent variable. The experimenter could have avoided this confounding if he had randomly assigned to the two treatment groups participants who signed up at different times, as in the first experiment described.



### Practice Thinking 4.2: Identifying Variables in Experiments

For each research example below, write the independent (or manipulated) variable and the dependent (or measured) variable. Think of an extraneous variable that is not controlled and that might confound the results.

1. An experimenter wanted to find out if a study strategy involving putting list items into categories was more effective for learning a list of words than a strategy involving rote rehearsal. One group of 30 randomly assigned participants studied words such as *chair*, *dog*, *rose*, *table*, *rabbit*, and *lily* by putting them into categories, such as “furniture,” “flowers,” and “pets.” The rote rehearsal group of 30 participants simply repeated the list of words over and over to themselves. Then both groups recalled the words one week later.

Independent variable: \_\_\_\_\_

Dependent variable: \_\_\_\_\_

Extraneous variable: \_\_\_\_\_

2. An experiment by Darley and Latané (1968) found that a person is more likely to help if alone than if part of a group. To replicate this study in a natural setting, Angie tested students at her university student center. An assistant would drop a stack of books in the presence of individual students walking alone or in front of groups of students walking together. Then Angie kept track of whether or not a subject (or subjects) helped in each case.

Independent variable: \_\_\_\_\_

Dependent variable: \_\_\_\_\_

Extraneous variable: \_\_\_\_\_

**Quasi-experiments** resemble true experiments in that they often involve the comparison of groups that undergo different treatment conditions; but unlike with true experiments, there is no true manipulation of an independent variable in quasi-experiments. Participants are not randomly assigned to treatment groups the way they are in true experiments. Instead, they are selected, often on the basis of some preexisting characteristic. For example, we might compare a group of college students with a group of high school students on their dating behaviors. Or we might compare men and women on their willingness to help. The comparison of college to high school students and of men to women does not entail true manipulation of the variables, even though they form different groups. We have simply selected people of different ages and sexes to be in our groups.

The problem is that, for instance, males and females might already differ from each other at the beginning of our study on a number of variables related to their willingness to help. Because we have merely selected males and females and cannot randomly assign participants to be male and female in a quasi-experiment, we are unable to control differences in our subjects that could be controlled through random assignment in a true experiment. Moreover, without truly manipulating an independent variable, we cannot establish time order; so we are not able to draw causal inferences from quasi-experiments, as we can with true experiments. Like true experiments, quasi-experiments can sometimes allow for control of extraneous variables—as when we test our groups under similar conditions in the laboratory—but possible preexisting differences in participants related to the sex variable remain uncontrolled.

In summary, the manipulation of independent variables allows the experimenter to meet the criteria of covariation and time order, and the control of extraneous variables allows for meeting the criterion of the elimination of plausible alternatives. At least with regard to making a causal inference, therefore, the experimental method provides better-quality data than the case study, correlational study, or quasi-experimental study. Table 4.3 summarizes the strengths and weaknesses of the various research designs we have discussed, with implications for the quality of data and evidence each provides.

Nevertheless, the conclusions based on scientific research are only as good as the quality of the evidence they are based on. Therefore, if a scientist did a research study and was not really measuring what was intended or perhaps made errors in measurement, then conclusions based on that research data could be erroneous. Fortunately, science is self-correcting, and the erroneous conclusion of the first scientist could be discovered by other scientists seeking to replicate and make sense of the observations of the first research study.

Table 4.3 also implies that we should be particularly persuaded when high-quality scientific research studies are used as evidence. When using scientific research as evidence in arguments, authors often cite the source author(s) and year of publication and mention the kind of research study that was done. Table 4.3 makes it clear that results of certain types of studies, such as true experiments, generally provide stronger support for a claim than do other types, such as case studies or other nonexperimental study designs. Scientific research is also used to support claims when a scientific authority or expert is cited, often someone who has written a literature review, summarizing the results of several studies supporting some hypothesis or theory.

These two citation methods demonstrate good practices in using scientific research as evidence, but we often hear arguments in the media and everyday life that do not specifically cite the study or research being discussed. For example, news reports will say, “Research shows . . .” or “Studies show . . .” without documenting the scientific research evidence being referred to. Although they have made a basic argument, their failing to cite specific research has weakened the argument. It also discourages critical thinking (CT) because it is harder to examine the quality of the evidence when no source is cited.

**TABLE 4.3** Strengths and Weaknesses of Scientific Research  
Methods/Designs Used as Sources of Evidence

Method/Design	Strengths	Weaknesses
<b>Case study</b> Detailed description of one or a few subjects	<ul style="list-style-type: none"> <li>Provides much information about one person</li> <li>May inform about a person with special or rare abilities, knowledge, or characteristics</li> </ul>	<ul style="list-style-type: none"> <li>May be unique and hard to replicate</li> <li>May not generalize to other people</li> <li>Cannot show cause and effect</li> </ul>
<b>Naturalistic observation</b> Observations of behavior made in the field or natural environment	<ul style="list-style-type: none"> <li>Allows observations to be readily generalized to the real world</li> <li>Can be a source of hypotheses</li> </ul>	<ul style="list-style-type: none"> <li>Allows little control of extraneous variables</li> <li>Cannot test treatments</li> <li>Cannot show cause and effect</li> </ul>
<b>Survey research</b> A method, often in the form of a questionnaire, that allows many questions to be asked	<ul style="list-style-type: none"> <li>Allows economical collection of much data</li> <li>Allows for study of many different questions at once</li> </ul>	<ul style="list-style-type: none"> <li>May have problems of self-reports, such as dishonesty, forgetting, and misrepresentation of self</li> <li>May involve biased sampling</li> </ul>
<b>Correlational study</b> A method for finding a quantitative relationship between variables	<ul style="list-style-type: none"> <li>Allows researcher to calculate the strength and direction of relation between variables</li> <li>Can be used to make predictions</li> </ul>	<ul style="list-style-type: none"> <li>Does not allow random assignment of participants or much control of subject variables</li> <li>Cannot test treatments</li> <li>Cannot show cause and effect</li> </ul>
<b>Quasi-experiment</b> A method for comparing treatment conditions without random assignment	<ul style="list-style-type: none"> <li>Allows comparison of treatments</li> <li>Allows some control of extraneous variables</li> </ul>	<ul style="list-style-type: none"> <li>Does not allow random assignment of participants or much control of subject variables</li> <li>Cannot show cause and effect</li> </ul>
<b>True experiment</b> A method for comparing treatment conditions in which variables can be controlled through random assignment	<ul style="list-style-type: none"> <li>Allows true manipulation of treatment conditions</li> <li>Allows random assignment and much control of extraneous variables</li> <li>Can show cause and effect</li> </ul>	<ul style="list-style-type: none"> <li>Cannot manipulate and test certain variables</li> <li>May control variables and conditions so much that they become artificial and unlike the “real world”</li> </ul>

Information from Bensley (2010) and Bensley (1998).

**Practice Thinking 4.3: Recognizing Kinds of Scientific Evidence**

In the space provided for each of the following examples, first identify the claim and the evidence supporting the argument being made. Then identify the kind of research study used as evidence. Finally, using Table 4.3, think of a possible limitation associated with that kind of evidence.

1. The primatologist Jane Goodall studied the interactions of chimpanzees from the same group with nongroup chimps. She observed that bands of male chimps of the same group would patrol the boundaries of their territories and sometimes gang up on and even kill other chimps that strayed into their territories. This research suggests that male chimpanzees engage in warlike behavior similar to that of human males.

Claim/evidence: \_\_\_\_\_

Kind of study: \_\_\_\_\_

Limitation: \_\_\_\_\_

2. After observing Anna O., both interviewing her and describing her behaviors, Freud argued that her numbness and other symptoms were due to her repressing her sexual feelings and that the conversion of those feelings to physical symptoms was a sort of unconscious defense mechanism.

Claim/evidence: \_\_\_\_\_

Kind of study: \_\_\_\_\_

Limitation: \_\_\_\_\_

3. Make your own argument about some hypothesis in psychology and support it with a type of evidence listed in Table 4.3.

Claim/evidence: \_\_\_\_\_

Kind of study: \_\_\_\_\_

Limitation: \_\_\_\_\_



### Practice Thinking 4.4: Analyzing Research for Causal Relations

Your task is to decide whether a conclusion based on the research results is warranted. Evaluate the quality of the evidence on the basis of the results presented and the research methods used. Only if all three criteria for causation are met should you conclude that a causal inference can be drawn from the study. To help you judge the quality of the evidence presented in each example, answer the questions that follow the study description.

An educational psychologist, investigating college students' use of study time, asked students to study the same material over a one-week period and measured the number of breaks each student took. At the end of the week, the students were tested on the material they learned. The psychologist found that the students who took more scheduled breaks also tended to score higher, whereas students who took fewer scheduled breaks tended to have lower test scores.

1. What kind of research method was used?

---

---

2. Which criterion/criteria for causation were met?

---

---

3. Can the psychologist accurately infer that study breaks cause better learning? Why or why not? (*Hint: Only true experiments allow causal inferences.*)

---

---

4. After hearing these results, Joe Student decided to take more study breaks the next time he studied. Is Joe's decision justified, given the evidence presented? Explain your answer.

---

---

5. If the psychologist in the study previously described could not accurately infer that study breaks result in better learning, which research method/design *would* put him in a better position to infer causation? Explain in detail.

---

---

True experiments put us in a better position to identify the causes of behaviors, but the debate does not end there. For instance, as discussed in previous chapters, many true experiments have supported social learning theories of aggression and the hypothesis that observing media violence leads to increases in aggression. Yet some researchers object that the true cause of aggression has not yet been identified because the true experiments conducted thus far do not yet have sufficiently high quality to draw this conclusion (Ferguson, 2015).

It is clear from the preceding discussion that drawing an appropriate conclusion from the research is complicated and that we must be careful not to make various thinking errors. It often requires identifying various types of nonscientific and scientific evidence and evaluating the quality and quantity of the evidence, as well as trying to identify relationships among variables. Only then can we finally draw a conclusion that is consistent with all the relevant evidence. Table 4.4 provides definitions and examples of thinking errors discussed in this chapter. Review these before applying your knowledge in Practice Thinking 4.5.

**TABLE 4.4** Summary of Thinking Errors Highlighted

Error Name	Description	How to Fix or Avoid It
Illusory correlation	Perceiving a correlation or association between two things when no correlation or association exists	Pay attention to cells B and C in a fourfold table, because people tend to focus mostly on cell A.
Confirmation bias	The tendency to attend to, seek, and give more weight to evidence that supports one's favored position rather than evidence that could disconfirm it	Consider the opposite or an alternate position—look for evidence that could disconfirm one's favored position.
Post hoc ("after this, therefore because of this") reasoning	Concluding after an event occurs that something that happened before it was the actual cause of the event	Don't assume that some action or situation that preceded another event was the actual cause of it; conduct a well-controlled experiment to see whether manipulating the first variable actually causes changes in the second.
Confusing correlation with causation	Believing that a variable that is simply covarying or correlated with another variable is its cause; or, less commonly, failing to infer causation from results of a well-controlled true experiment	Make sure the action or situation thought to be the cause of an event actually occurred first; look for other possible causes and see if they can be eliminated; conduct a true experiment and don't be fooled into thinking that correlations or quasi-experiments can show causation.



### Practice Thinking 4.5: Identifying and Fixing Thinking Errors

Identify the kind of thinking errors in the following examples and explain how to fix them.

1. A psychologist found that people who had been physically or sexually abused as children often tended to later become abusers themselves. He concluded that early abuse of a person causes that person to become abusive later.

a. Thinking error: \_\_\_\_\_

b. How to fix it: \_\_\_\_\_

2. Niki's mother was given a preliminary diagnosis that she had developed cancer. Niki had heard that maintaining a positive attitude could stave off cancer. She encouraged her mother to think positively about beating her cancer. In the next two weeks, Niki's mother said she felt better; when she went back to the doctor, he said he could find no trace of the cancer. Niki was convinced that positive thinking had rid her mother of cancer.

a. Thinking error: \_\_\_\_\_

b. How to fix it: \_\_\_\_\_

3. As Kevin took his first exam, he looked over each true–false question and thought about how confusing it was for him. A pattern seemed to emerge. He noticed that every time a question confused him, the answer seemed to be false. He was surprised when he got his test back and found that about an equal number of the confusing items were true as were false.

a. Thinking error: \_\_\_\_\_

b. How to fix it: \_\_\_\_\_

4. Julie was quite sure that her roommate Sarah was depressed. She asked whether Sarah ever felt a lack of energy, and Sarah answered that yes, she did. Another time, Julie asked if Sarah was unhappy with her classes, to which Sarah replied, "Sometimes." At a party with people they did not know, Sarah seemed nervous and anxious. Julie thought to herself, "Well, she lacks energy, is unhappy with her classes, and seems anxious. People who are depressed are often anxious, too. Sarah's depressed for sure."

a. Thinking error: \_\_\_\_\_

b. How to fix it: \_\_\_\_\_

### WHAT IS SCIENTIFIC THINKING?

As we have seen, scientific thinking is complex. We have discussed how scientists use inductive reasoning to generalize from research studies to justify theories, as when Pierre Paul Broca induced that a certain area of the brain regulated speech

production from observations of people with brain damage in the left frontal area who had speech-production problems. From theories, scientists deduce hypotheses, such as the prediction that a group of people with damage to Broca's area would have speech-production problems. Then they conduct research to test these hypotheses. If a prediction is confirmed, it lends inductive support to the theory; if it is not confirmed, then support for the theory is weakened. Although this description captures some of the scientific method that scientists use to develop their ideas, it does not address how they come up with those ideas nor how they think about specific problems. The following discussion introduces some of these issues, but the online supplement to this chapter goes into greater detail.

Taking a different approach, researchers in cognitive psychology have studied scientific thinking as a kind of problem solving (Dunbar & Fugelsang, 2005; Klahr, Matlen, & Jirout, 2013; Newell & Simon, 1972). Suppose you are striving to reach some goal but do not know how to get from your current state to the goal state—here you have a **problem**. For example, you are interested in meeting someone in your class, but you don't know how to meet this person without being too obvious. This presents a problem because you lack the knowledge and strategy to progress from your initial state to the goal of being introduced.

Likewise, scientists are problem solvers who start out not knowing how to achieve the goal state of solving some scientific problem. For instance, they may want to test a new hypothesis but lack the right method, equipment, or other resources to test it effectively. To solve a problem, scientists go through several stages: First, they must identify or find the problem; next, they must represent or understand the problem in order to generate a strategy that might lead to a solution; then they can apply the strategy to the problem and monitor whether they have solved it. It should therefore seem clear to you that scientific problem solving is closely related to CT because scientists reason about the quality of the information they have and the problem-solving strategies they use. They select and apply the best strategy that is practically available to them and then engage in metacognitive monitoring to evaluate their progress toward a solution (Willingham, 2007).

Often, scientific problem solving also involves **creative thinking**, which requires a scientist to think about the problem in new and useful ways. To solve difficult problems and questions, scientists must propose new hypotheses, invent new tools, and develop new ways to conduct their research. For example, the discovery of the double-stranded structure of DNA by James Watson and Francis Crick in 1953 required the development of new X-ray crystallography equipment; subsequent experiments by Maurice Wilkins and Rosalind Franklin made

use of that new equipment. It involved evaluating and interpreting the data to decide whether they best fit a two-stranded or three-stranded DNA molecule. This, in turn, required Wilkins and Franklin to represent the possibilities using different models, to weigh the evidence for each, and then to evaluate all the information to draw the best conclusion (Weisberg, 2006). Thus, creative scientific thinking is like problem solving, but it also resembles CT in a general way (Willingham, 2007).

## SUMMARY

In their study of the natural world, psychologists and other scientists seek to identify relationships between variables. These relationships are either associations or causal relationships. In associations (correlations), two variables simply change together in a consistent way. People seem to have a natural tendency to seek and find relationships. Unfortunately, they often find relationships that aren't really there, demonstrating the thinking error of illusory correlation.

This occurs because they pay attention to the co-occurrences of two variables, such as observing when the moon is full and people behave abnormally but not systematically observing when the moon is full and people are *not* behaving abnormally—or, conversely, disregarding abnormal behavior when the moon is *not* full. Once someone accepts this illusory correlation, he or she may only look for evidence to support his or her belief or to minimize the evidence that does not support it, demonstrating the thinking error of confirmation bias. In contrast, scientists who calculate correlations, conduct experiments, and review other people's research take into account negative evidence that does not support a relationship. In this way, science becomes self-correcting, allowing researchers to discover which relationships do *not* exist—a valuable and informative practice that allows them to correct misconceptions.

A causal relationship is harder to show than an association. In causal relationships, one variable—the cause—must occur before a second variable—the effect—and must lead to consistent changes in that second variable. Showing causation also requires eliminating any third or other variable that could account for the resulting changes. Only one research design, the true experiment, puts the researcher in a position to show cause and effect. To produce time order, the experimenter manipulates an independent variable (the cause) so that it precedes a dependent variable (the effect). She uses probability and statistics to determine

whether the values of the two variables are co-varying, that is, if they change together consistently. To eliminate plausible alternative explanations (third or extraneous variables that could affect the dependent variable), she uses control procedures, such as random assignment of participants to treatment groups.

The degree to which different research methods can show causation is important to the quality of the evidence each can provide. For example, case studies, survey research, correlational studies, and field studies simply measure variables without any manipulation of them and so cannot show time order. Although quasi-experiments resemble true experiments, they—like the other nonexperimental methods—do not use random assignment to treatment groups and show less control over extraneous variables. Because only the true experiment meets the criteria for showing causation, it provides the strongest evidence. Failing to realize this, people often mistakenly infer causation when only a correlation exists, or they mistakenly assume that a variable (action or situation) that simply occurs prior to a second variable (event) actually caused the change in the second event (the post hoc fallacy).

In addition to the quality of evidence provided, the quantity of evidence affects the strength of an inductive argument. For instance, a case study is conducted on only one or a few people and so provides little in the way of quantitative support. In general, the larger the sample, the stronger the support a study can offer. Likewise, a theory is more strongly supported when many studies support it rather than fewer; however, quality trumps quantity. A poorly executed study with a large sample offers weak support.

Scientific thinking involves the use of inductive reasoning to generalize from research evidence to justify hypotheses and theories, and then deducing predictions from those theories to test. But it involves other kinds of thinking, too—such as problem solving and creative thinking, both of which resemble CT in important ways.



#### **Practice Thinking 4.6: WHAT DO YOU THINK NOW?**

Please explain how you know.

1. When people join a cult, do they accept all the cult's beliefs and lose their ability to think critically? How could you find out?

2. Does observing aggressive behaviors in movies, TV, or video games cause a person to behave aggressively? How could you find out?

---

3. Does having been sexually abused as a child lead to personality problems as an adult and cause the victim to later become a sexual abuser?

---

4. Can science answer all questions? Why or why not?

---

---

## REVIEW QUESTIONS

---

1. What are some differences between scientific and nonscientific sources of evidence?
2. What is science? What are its advantages as an approach to knowledge?
3. What are the three criteria for causation?
  - Why does the example of Lee Fried's presumed precognition not demonstrate causation?
  - How is the quality of scientific evidence related to causality?
4. Compare and contrast common research methods with respect to their strengths and weaknesses, especially with regard to showing causation.
  - What is survey research? What are its strengths and weaknesses?
  - What is a field study? What are its strengths and weaknesses?
  - What is correlation? What are its strengths and weaknesses?
  - What is a true experiment? What are its strengths and weaknesses?
  - What is a quasi-experiment? What are its strengths and weaknesses?
  - Can you identify independent, dependent, and extraneous variables?
  - What are confounding variables and how do they relate to internal validity?
  - Can you evaluate a research study to determine whether it allows for a causal inference?

5. What thinking errors are featured in this chapter?
  - What is an example of each?
  - How do you correct or counter each one?
6. What is scientific thinking? Is it just induction and deduction?
  - How does scientific thinking involve problem solving and creative thinking?
  - How is CT related to scientific thinking? To scientific problem solving? To scientific creativity?
7. What psychological misconceptions are featured in this chapter? Explain why each is a misconception.

