2

CHAPTER TWO

Empiricism: How to Know Things

The Science of Observation: Saying What

CULTURE AND COMMUNITY Expecting a Helping Hand? It Depends Where You Are

Measurement

Samples

Demand Characteristics

The Blind Observer

The Science of Explanation: Saying Why

Correlation and Causation

Matched Samples and Matched Pairs

HOT SCIENCE Establishing Causality in the Brain

Experimentation

Drawing Conclusions

The Ethics of Science: Saying Please and Thank You

WHERE DO YOU STAND? The Morality of Immoral Experiments

The Methods of Psychology

LORI AND REBA SCHAPPELL ARE HAPPY to be twins. One is an award-winning country music singer; one is a wisecracking hospital worker who likes strawberry daiquiris. Despite their different interests and different temperaments, they get along quite well and love each other dearly. That's a good thing because Lori and Reba share more than the same parents and the same birthday. Lori and Reba are conjoined twins who have been attached at the forehead since birth. When asked whether they would ever consider being surgically separated, Reba seems perplexed: "Our point of view is no, straight-out no. You'd be ruining two lives in the process" (Angier, 1997). If you find this hard to believe, then welcome to the club. Conjoined twins are routinely separated at birth even when this means crippling both or killing one of them because surgeons and parents—like most of us—can't imagine that a conjoined life is really worth living. And yet, conjoined twins don't seem to share that view. As one medical historian noted, "The desire to remain together is so widespread among communicating conjoined twins as to be practically universal. . . .I have yet to find an instance in which conjoined twins have sought out separation" (Dreger, 1998).

Are conjoined twins really as happy as they claim, or are they simply fooling themselves? Do parents and doctors have the right to impose dangerous surgery on infants who would otherwise grow up to refuse it? Such questions have moral, religious, and philosophical answers, but they can have scientific answers as well. If we could find some way to measure a psychological property such as happiness, then we could use scientific methods to determine who has it and who doesn't and to discover what kinds of lives promote or preclude it. Is a conjoined life a wonderful life, or is it society's responsibility to separate conjoined twins whenever possible? As you are about to see, psychological methods are designed to provide answers to questions like this one. ■

AP PHOTO/READING EAGLE, JOHN A. SECOGES

Are conjoined twins less happy than singletons? Reba (left) and Lori (right) Schappell are sisters who say that the answer is no.

Empiricism: How to Know Things

When ancient Greeks sprained their ankles, caught the flu, or accidentally set their togas on fire, they had to choose between two kinds of doctors: dogmatists (from *dogmatikos,* meaning "belief"), who thought that the best way to understand illness was to develop theories about the body's functions, and empiricists (from *empeirikos,* meaning "experience"), who thought that the best way to understand illness was to observe sick people. The rivalry between these two schools of medicine didn't last long, however, because the people who chose to see dogmatists tended to die, which wasn't very good for repeat business. It is little wonder that today we use the word *dogmatism* to describe the tendency for people to cling to their assumptions and the word **empiricism** to describe *the belief that accurate knowledge of the world requires observation of it*. The fact that we can answer questions about the world by observation may seem obvious to you, but this obvious fact is actually a relatively new discovery. Throughout most of human history, people have trusted authority to answer important questions about the world, and it is only in the last millennium (and especially in the past three centuries) that people have begun to trust their eyes and ears more than their elders. Empiricism has proved to be a profitable approach to understanding natural phenomena, but using this approach requires a **method**, which is *a set of rules and techniques for observation that allow observers to avoid the illusions, mistakes, and erroneous conclusions that simple observation can produce.*

"Are you just pissing and moaning, or can you verify what you're saying with data?"

● Why is it so hard to study people scientifically?

Human behavior is relatively easy to observe, so you might expect psychology's methods to be relatively simple. In fact, the empirical challenges facing psychologists are among the most daunting in all of modern science, and thus psychological methods are among the most sophisticated. Three things make people especially difficult to study:

- *Complexity:* Psychologists study the single most complex object in the known universe. No galaxy, particle, molecule, or machine is as complicated as the human brain. Scientists can barely begin to say how the 500 million interconnected neurons that constitute the brain give rise to the thoughts, feelings, and actions that are psychology's core concerns.
- *Variability:* In almost all the ways that matter, one *E. coli* bacterium is pretty much like another. But people are as varied as their fingerprints. No two individuals ever do, say, think, or feel exactly the same thing under exactly the same circumstances.
- *Reactivity:* An atom of cesium-133 oscillates 9,192,631,770 times per second regardless of who's watching. But people often think, feel, and act one way when they are being observed and a different way when they are not.

In short, human beings are tremendously complex, endlessly variable, and uniquely reactive, and these attributes present a major challenge to the scientific study of their behavior. As you'll see, psychologists have developed a variety of methods that are designed to meet these challenges head-on.

● *The fish are probably annoyed by the wide variety of people who invade their territory on Labor Day, but hey . . . life's a beach.*

The Science of Observation: Saying What

There is no escaping the fact that you have to observe *what* people do before you can try to explain *why* they do it. To *observe* something means to use your senses to learn about its properties. For example, when you observe a round, red apple, your brain is using the pattern of light that is falling on your eyes to draw an inference about the apple's identity, shape, and color. That kind of informal observation is fine for buying fruit but not for doing science. Why? First, casual observations are notoriously unstable. The same apple may appear red in the daylight and crimson at night or spherical to one person and elliptical to another. Second, casual observations can't tell us about many of the properties in which we might be interested. No matter how long and hard you look, you will never be able to discern an apple's crunchiness or pectin content simply by watching it. If you want to know about those properties, you must do more than observe. You must *measure*.

Measurement

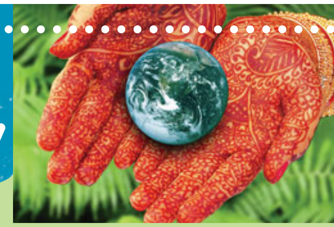
You probably think you know what *length* is. But if you try to define it without using the word *long*, you get tongue-tied pretty quickly. We use words such as *weight, speed,* or *length* all the time in ordinary conversation without realizing that each of these terms has an **operational definition,** which is a *description of a property in measurable terms*. For example, the operational definition of the property we casually refer to as *length* is "the change in the location of light over time." When we say that a bookshelf is "a meter in length," we are actually saying how long it takes a particle of light to travel from one end of the shelf to the other. (In case you're interested, the answer is 1/299,792,458th of a second.) Operational definitions specify the concrete events that count as instances of an abstract property. The first step in making any measurement is to define the property we want to measure in concrete terms.

The second step is to find a way to detect the concrete terms that our definition describes. To do this we must use a **measure**, which is *a device that can detect the events to which an operational definition refers*. For example, length is the change in the location of light over time, and we can detect such changes by using a photon detector (which tells us the location of a particle of light) and a clock (which tells us how long it took the particle of light to travel from one location to another). Once we have determined just how far a photon travels in 1/299,792,458th of a second, we can make our next measurement a lot less expensive by marking that distance on a piece of wood and calling it a ruler.

Defining and *detecting* are the two tasks that allow us to measure physical properties, and these same two tasks allow us to measure psychological properties as well. If we wanted to measure Lori Schappell's happiness, for example, our first task would be to develop an operational definition of that property—that is, to specify some concrete, measurable event that will count as an instance of happiness. For example, we might define happiness as the simultaneous contraction of the *zygomatic major* (which is the muscle that makes your mouth turn up when you smile) and the *orbicularis oculi* (which is the muscle that makes your eyes crinkle when you smile). After defining happiness as a specific set of muscular contractions, we would then need to measure those contractions, and the **electromyograph (EMG)**—which is *a device that measures muscle contractions under the surface of a person's skin*—would do splendidly. Once we have defined happiness and found a way to detect the concrete events that our definition supplies, we are in a position to measure it (FIGURE 2.1).

● **How could you measure happiness?**

Culture & Community



Expecting a Helping Hand? It Depends Where You Are

Robert Levine of California State University–Fresno sent his students to 23 large international cities for an observational study in the field. Their task was to observe helping behaviors in a naturalistic context. In two versions of the experiment, students pretended to be either blind or injured while trying to cross a street, while another student stood by to observe whether anyone would come to help. A third version involved a student dropping a pen to see if anyone would pick it up.

RICARDO FUNARI/BRAZIL PHOTOS/ALAMY

The results showed that people helped in all three events fairly evenly within cities, but there was a wide range of response between cities. Rio de Janeiro, Brazil, came out on top as the most helpful city in the study with an overall helping score of 93%. Kuala Lampur, Malaysia, came in last with a score of 40%, while New York City placed next to last with a score of 45%. On average, Latin American cities ranked most helpful (Levine, Norenzayan, & Philbrick, 2001).

empiricism Originally a Greek school of medicine that stressed the importance of observation, and now generally used to describe any attempt to acquire knowledge by observing objects or events.

method A set of rules and techniques for observation that allow researchers to avoid the illusions, mistakes, and erroneous conclusions that simple observation can produce.

operational definition A description of an abstract property in terms of a concrete condition that can be measured.

measure A device that can detect the measurable events to which an operational definition refers.

electromyograph (EMG) A device that measures muscle contractions under the surface of a person's skin.

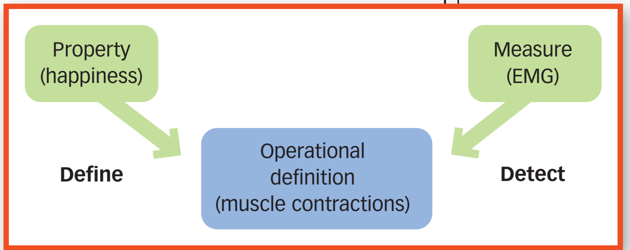


FIGURE 2.1
Sources of Invalidity. *The process of defining links properties to operational definitions, and the process of detecting links operational definitions to measures. Invalidity can result from problems in either of these links.*

"Are you (a) contented, (b) happy, (c) very happy, (d) wildly happy, (e) deliriously happy?"

But is this the *right* way to measure happiness? That's hard to say. There are many ways to define the same property and many ways to detect the events that this definition supplies. For instance, we could detect the muscular contractions involved in smiling by using EMG, or we could detect them by asking a human observer to watch a participant's face and tell us how often the participant smiled. We could even define happiness as a person's self-assessment of his or her own emotional state, in which case we could measure it by asking people how happy they feel and recording their answers. With so many options for defining and detecting happiness, how are we to choose among them?

The best kinds of measurements share three properties: validity, reliability, and power. **Validity** is *the characteristic of an observation that allows one to draw accurate inferences from it*. There are two kinds of validity. First, the operational definition must adequately define the property ("construct validity"). In other words, the distance that a proton travels is a good way to define length, but it's not a valid way to define happiness. Second, the operational definition must be related to other operational definitions of the same property ("predictive validity"). In other words, if an operational definition such as smiling is linked to a property such as happiness, then it should also be linked to other operational definitions of the same property—such as a person's likelihood of saying, "I sure am happy right now."

The other important property of a good measurement is **reliability**, which is *the tendency for a measure to produce the same result whenever it is used to measure the same thing*. For example, if a person's zygomatic muscle did not move for 10 minutes, we would expect the EMG to produce the same reading for 10 minutes. If the EMG produced different readings from one minute to the next, then it would be an unreliable measure that was detecting differences that weren't really there. A good measure must be reliable. The flip side of reliability is **power**, which is *the tendency for a measure to produce different results when it is used to measure different things*. If a person's zygomatic muscle moved continuously for 10 minutes, we would expect the EMG to produce different readings in those 10 minutes. If the EMG instead produced the same reading from one minute to the next, then it would be a weak or powerless measure that was failing to detect differences that were really there. Reliable and powerful measures are those that detect the conditions specified by an operational definition (a) when they happen and (b) *only* when they happen.

Validity, reliability, and power are prerequisites for accurate measurement. But once you've got a good ruler in hand, the next step is to find something to measure with it. Psychologists have developed techniques for doing that, too.

MEDIABAKERY

COURTESY SARTORIUS CORPORATION

A bathroom scale and a laboratory balance both measure weight, but the balance is more likely to provide exactly the same measurement when it is used to weigh the same object twice (reliability) and more likely to provide different measurements when it is used to weigh two objects that differ by just a fraction of a gram (power). Not surprisingly, the bathroom scale sells for around $30 and the balance for around $3,000. Power and reliability don't come cheap.

Samples

If a pig flew over the White House, it wouldn't matter whether other pigs could do the same trick. The fact that just one pig flew just one time would challenge our most cherished assumptions about animal physiology, aerodynamics, and national security and would thus be an observation well worth making. Similarly, individuals sometimes do remarkable things that deserve close study, and when psychologists study them closely, they are using the **case method**, which is *a method of gathering scientific knowledge by studying a single individual*. For example, the physician Oliver Sacks described his observations of a brain-damaged patient in a book titled *The Man Who Mistook His Wife for a Hat,* and those observations were worth making because this is a rather unusual mistake for a man to make. As you saw in Chapter 1, people with unusual abilities, unusual experiences, or unusual deficits often provide important insights about human psychology.

● How can an exceptional case teach us about normal behavior?

But exceptional cases are the exception, and more often than not, psychologists are in the business of observing *un*exceptional people and trying to explain why they think, feel, and act as they do. Of course, it's not possible to observe every ordinary person in the world. Even if we consider a subset (e.g., every pair of conjoined twins currently alive), it's still not possible to observe everyone who meets those criteria—if only because these people may be spread across many different continents. Instead of observing the entire **population**, *the complete collection of objects or events that might be measured,* psychologists observe a **sample**, *a partial collection of objects or events that is measured.* If the sample is relatively large and well chosen, then the behavior of individuals in the sample should be representative of the larger population. (For more on sampling techniques, see the appendix.)

"This fundamentally changes everything we know about elephants!"

Demand Characteristics

Once psychologists have settled on a valid and reliable measurement, and constructed a representative sample to study, the next problem is to figure out how to apply that measurement to the sample in the most accurate way possible. One problem is that, while psychologists are trying to discover how people really *do* behave, people are often trying to behave as they think they *should* behave. People pick their noses, exceed the speed limit, read each other's mail, and skip over major sections of *War and Peace*, and they are especially likely to do these things when they think no one is looking. They are much less likely to indulge in these behaviors if they are aware that someone is observing them and taking notes. **Demand characteristics** are *those aspects of a setting that cause people to behave as they think an observer wants or expects them to behave*. They are called demand characteristics because they seem to "demand" or require that people say and do things that they normally might not. If you have ever been asked the question "Do you think these jeans make me look fat?," then you have experienced a demand characteristic. Demand characteristics hinder our attempts to measure behavior as it normally unfolds.

● Why do people act differently when they know they're being observed?

One way psychologists try to avoid this problem is to observe people without their knowledge. **Naturalistic observation** is *a technique for gathering scientific knowledge by unobtrusively observing people in their natural environments*. For example, naturalistic observation reveals that the biggest groups tend to leave the smallest tips in restaurants (Freeman et al., 1975), that hungry shoppers buy the most impulse items at the grocery store (Gilbert, Gill, & Wilson, 2002), and that Olympic athletes smile more when they win the bronze rather than the silver medal (Medvec, Madey, & Gilovich, 1995). All of these conclusions are the result of measurements made by psychologists who observed people who didn't know they were being observed. It is unlikely that any of these things would have

ONLY HUMAN

MAYBE THEY COULD PASS A LAW OF LARGE NUMBERS? In 1997, David Cook of Caledonian University in Glasgow, Scotland, told the British Psychological Society's annual conference that his 3-year study shows that politicians have significant behavior patterns in common with criminal psychopaths. Cook said that criminals were relatively easy to analyze but that he did not have as much data as he would like on politicians. "They don't like to be studied," he said.

validity The characteristic of an observation that allows one to draw accurate inferences from it.

reliability The tendency for a measure to produce the same result whenever it is used to measure the same thing.

power The tendency for a measure to produce different results when it is used to measure different things.

case method A method of gathering scientific knowledge by studying a single individual.

population The complete collection of participants who might possibly be measured.

sample The partial collection of people who actually were measured in a study.

demand characteristics Those aspects of an observational setting that cause people to behave as they think an observer wants or expects them to behave.

naturalistic observation A method of gathering scientific knowledge by unobtrusively observing people in their natural environments.

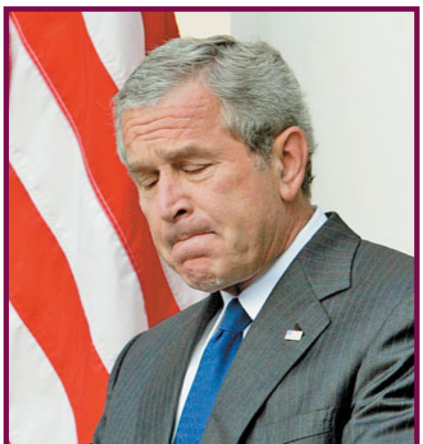
DANIEL PEEBLES PHOTO

This bar on 10th Avenue in New York City has a "one-way" mirror in its unisex restroom. Customers see their reflections in the restroom's mirror, and people who are walking down the street see the customers. Are the customers influenced by the fact that pedestrians may be watching them? Hard to say, but one observer did notice a suspiciously "high percentage of people who wash their hands" (Wolf, 2003).

happened in exactly the same way if the diners, shoppers, and athletes had known that they were being scrutinized.

Unfortunately, there are two reasons why naturalistic observation cannot by itself solve the problem of demand characteristics. First, some of the things psychologists want to observe simply don't occur naturally. For example, if we wanted to know whether people who have undergone sensory deprivation perform poorly on motor tasks, we would have to hang around the shopping mall for a very long time before a few dozen blindfolded people with earplugs just happened to wander by and start typing. Second, some of the things that psychologists want to observe can only be gathered from direct interaction with a person—for example, by administering a survey, giving tests, conducting an interview, or hooking someone up to an EEG. If we wanted to know how often people worried about dying, how accurately they could remember their high school graduation, how quickly they could solve a logic puzzle, or how much electrical activity their brain produced when they felt happy, then simply observing them would not do the trick.

When psychologists cannot avoid demand characteristics by hiding in the bushes, they often avoid them by hiding other things instead. For instance, people are less likely to be influenced by demand characteristics when they cannot be identified as the originators of their actions, and psychologists often take advantage of this fact by allowing people to respond privately (e.g., by having them complete questionnaires when they are alone) or anonymously (e.g., by failing to collect personal information, such as the person's name or address). Another technique that psychologists use to avoid demand characteristics is to measure behaviors that are not susceptible to demand. For instance, behaviors can't be influenced by demand characteristics if they aren't under voluntary control. You may not want a psychologist to know that you are feeling excited, but you can't prevent your pupils from dilating when you feel aroused. Behaviors are also unlikely to be influenced by demand characteristics when people don't know that the demand and the behavior are related. You may want a psychologist to believe that you are concentrating on a task, but you probably don't know that your blink rate slows when you are concentrating and thus you won't fake a slow blink.



WIN MCNAMEE/GETTY IMAGES

One way to avoid demand characteristics is to measure behaviors that people are unable or unlikely to control, such as facial expressions, reaction times, eye blink rate, and so on. For example, when people feel anxious, they tend to involuntarily compress their lips, as President George W. Bush did in this 2006 photo taken as he gave a speech in the Rose Garden.

All of these tricks of the trade are useful, of course, but the very best way to avoid demand characteristics is to keep the people who are being observed (known as *participants*) from knowing the true purpose of the observation. When participants are kept "blind" to the observer's expectations—that is, when they do not know what the observer expects them to do—then they cannot strive to meet those expectations. If you did not know that a psychologist was studying the effects of baroque music on mood, then you would not feel compelled to smile when the psychologist played Bach's *Air on*

G String. This is why psychologists often do not reveal the true purpose of a study to the participants until the study is over.

Of course, people are clever and curious, and when psychologists don't tell them the purpose of their observations, participants generally try to figure it out for themselves ("I wonder why the psychologist is playing the violin and watching me"). That's why psychologists sometimes use *cover stories,* or misleading explanations that are meant to keep participants from discerning the true purpose of an observation. For example, if a psychologist wanted to know how baroque music influenced your mood, he or she might tell you that the purpose of the study was to determine how quickly people can do logic puzzles while music plays in the background. (We will discuss the ethical implications of deceiving people later in this chapter.) In addition, the psychologist might use *filler items,* or pointless measures that are meant to mask the true purpose of the observation. So, for example, he or she might ask you a few questions that are relevant to the study ("How happy are you right now?") and a few that are not ("Do you like cats more or less than dogs?"), which would make it difficult for you to guess the purpose of the study from the nature of the questions you were asked. These are just a few of the techniques that psychologists use to avoid demand characteristics.

● Why is it sometimes important that participants not be aware of an experiment's true purpose?

double-blind observation An observation whose true purpose is hidden from the researcher as well as from the participant.

The Blind Observer

Participants aren't the only ones whose behavior can interfere with valid and reliable measurement. The behavior of the observers can interfere, too. After all, observers are human beings, and like all human beings, they tend to see what they expect to see. This fact was demonstrated in a classic study in which a group of psychology students were asked to measure the speed with which a rat learned to run through a maze (Rosenthal & Fode, 1963). Some students were told that their rat had been specially bred to be "maze dull" (i.e., slow to learn a maze), and others were told that their rat had been specially bred to be "maze bright" (i.e., quick to learn a maze). Although all the rats were actually the same breed, the students who *thought* they were measuring the speed of a dull rat reported that their rats took longer to learn the maze than did the students who *thought* they were measuring the speed of a bright rat. In other words, the rats seemed to do just what the students who observed them expected them to do.

Why did this happen? *First, expectations can influence observations*. It is easy to make errors when measuring the speed of a rat, and expectations often determine the kinds of errors people make. Does putting one paw over the finish line count as "learning the maze"? If the rat falls asleep, should the stopwatch be left running or should the rat be awakened and given a second chance? If a rat runs a maze in 18.5 seconds, should that number be rounded up or rounded down before it is recorded in the log book? The answers to these questions may depend on whether one thinks the rat is bright or dull. The students who timed the rats probably tried to be honest, vigilant, fair, and objective, but their expectations influenced their observations in subtle ways that they could neither detect nor control. Second, *expectations can influence reality.* Students who expected their rats to learn quickly may have unknowingly done things to help that learning along—for example, by muttering, "Oh, no!" when the bright rat turned the wrong way in the maze or by petting the bright rat more affectionately than the dull rat and so on. (We shall discuss these phenomena more extensively in Chapter 16.)

● When might a computer run a better experiment than a human being?

Observers' expectations, then, can have a powerful influence on both their observations and on the behavior of those whom they observe. Psychologists use many techniques to avoid these influences, and one of the most common is the **double-blind observation,** which is *an observation whose true purpose is hidden from both the observer and the participant*. For example, if the students had not been told which rats were bright and which were dull,

AP PHOTO/TIMES RECORD, KAIA LARSEN

People's expectations can influence their observations. On September 10, 2002, Gurdeep Wander boarded an airplane with three other dark-skinned men who had no luggage, switched seats, and got up several times to use the restroom. This was enough to convince the pilot to make an emergency landing in Arkansas and have Mr. Wander arrested as a potential terrorist. Mr. Wander is an American citizen who works at Exxon and was on his way to a convention.

GABRIEL BOUYS/AFP/GETTY IMAGES

The New York Times

STOCK PRICES SLUMP $14,000,000,000 IN NATION-WIDE STAMPEDE TO UNLOAD; BANKERS TO SUPPORT MARKET TODAY

PREMIER ISSUES HARD HIT

Unexpected Torrent of Liquidation Again Rocks Markets.

DAY'S SALES 9,212,800

Nearly 3,000,000 Shares Are Traded In Final Hour—The Tickers Lag 167 Minutes.

NEW RALLY SOON BROKEN

Selling by Europeans and "Mob Psychology" Big Factors in Second Big Break.

The second hurricane of liquidation within four days hit the stock market yesterday. It came suddenly, and violently, after holders of stocks had been lulled into a sense of security by the rallies of Friday and Saturday. It was a country-wide collapse of open-market security values in which the declines established and the actual losses taken in dollars and cents were probably the most disastrous and far-reaching in the history of the Stock Exchange.

That the storm has now blown itself out, that there will be organized support to put an end to a reaction which has ripped billions of dollars from market values, appeared certain last night from statements by leading bankers.

Although total estimates of the losses on securities are difficult to make, because of the large number of them not listed on any exchange, it was calculated last night that the total shrinkage in American securities on all exchanges yesterday had aggregated some $14,000,000,000, with a decline of about $10,000,000,000 in New York Stock Exchange securities. The figure is necessarily a rough one, but nevertheless gives an idea of the dollars and cents recessions in one of the most extraordinary declines in the history of American markets.

© THE NEW YORK TIMES

Worried customers lined up in 2008 after the IndyMac bank failed (above). Only the intervention of federal government prevented a repeat of the events of 1929. People's expectations can cause the phenomena they expect. In 1929, investors who expected the stock market to collapse sold their stocks and thereby caused the very crisis they feared. The New York Times *attributed the crash to "mob psychology" (right).*

then they would not have *had* any expectations about their rats. It is common practice in psychology to keep the observers as blind as the participants. For example, measurements are often made by research assistants who do not know what a particular participant is expected to say or do and who only learn about the nature of the study when it is concluded. Indeed, many modern studies are carried out by the world's blindest experimenter: a computer, which presents information to participants and measures their responses without any expectations whatsoever.

summary quiz [2.1]

1. The belief that accurate knowledge of the world requires observation of it is called

a. measurement.
b. empiricism.
c. validity.
d. naturalistic observation.

2. A set of rules and techniques for observation necessary to avoid mistakes that simple observation can produce is called

a. a method.
b. a measure.
c. an operational definition.
d. empiricism.

3. A device that can detect the event to which an operational definition refers is called

a. empiricism.
b. the case method.
c. a measure.
d. a detector.

4. Professor Craig developed a new test that supposedly measured IQ. When many individuals were given this test on two separate occasions, their scores showed little consistency from the first testing to the second. Professor Craig's test apparently lacked

a. validity.
b. reliability.
c. power.
d. demand characteristics.

correlation The "co-relationship" or pattern of covariation between two variables, each of which has been measured several times.

variable A property whose value can vary or change.

The Science of Explanation: Saying Why

The techniques discussed so far allow us to construct valid, reliable, powerful, and unbiased measures of properties such as happiness; to use those instruments to measure the happiness of a sample without demand characteristics; and to draw conclusions about the happiness of a population. Although scientific research always begins with the careful measurement of properties, its ultimate goal is typically the discovery of *causal relationships between properties*. We may find that happy people are more altruistic than unhappy people, but what we really want to know is whether their happiness is the cause of their altruism. Measurements can tell us how *much* happiness and altruism occur in a particular sample, but they cannot tell us (a) whether these properties are related and, (b) if so, whether their relationship is causal. As you will see, scientists have developed some clever ways of using measurement to answer these questions.

Correlation and Causation

If you wanted to determine whether there is any sort of relationship between happiness and altruism, you'd have to collect some data. (You are an empiricist, after all). You might, for example, measure the happiness and altruism of a few dozen people and make a table like the one shown in **TABLE 2.1**. Inspecting the table, you would quickly notice that altruism and happiness tend to occur together far more often than not. When two properties occur together, often we say they are **correlated**, which means that *the value of one is systematically related to the value of the other.* When the value of altruism is high, then the value of happiness tends to be high, too. *Correlation* is short for *co-relationship*.

● Televised violence and aggression are correlated. Does that mean televised violence causes aggressiveness?

As you look at Table 2.1, you might be tempted to conclude that the reason why happiness and altruism are correlated is that happiness causes altruism. You should resist that temptation. The fact that two **variables**, *properties whose values can vary across individuals or over time,* such as altruism and happiness are correlated does not mean that one necessarily causes the other.

Consider an example. Many studies (see Huesmann et al., 2003) have found a positive correlation between the amount of violence a child sees on television (let's call this variable *X*) and the aggressiveness of the child's behavior (let's call this variable *Y*). The more violence a child sees, the more aggressive that child is likely to be. But does that mean that seeing violence *causes* aggression? Not necessarily. It may be that watching violence on TV (*X*) causes aggressiveness (*Y*), but it may also be that aggressiveness (*Y*) causes children to watch televised violence (*X*). For example, children who are naturally aggressive may enjoy televised violence more than those who aren't and therefore may seek opportunities to watch it.

To make matters more complicated, it could even be the case that a third variable (*Z*) causes children both to be aggressive (*Y*)

TABLE 2.1

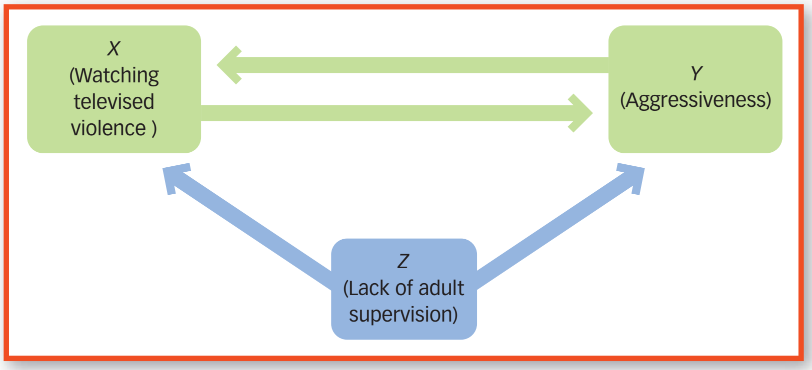
Hypothetical Data of the Relationship between Happiness and Altruism

Participant	Happiness	Level of Altruism
1	Happy	High
2	Happy	High
3	Unhappy	Low
4	Unhappy	Low
5	Happy	High
6	Happy	High
7	Unhappy	Low
8	Unhappy	Low
9	Happy	High
10	Happy	High
11	Unhappy	Low
12	Unhappy	Low
13	Happy	High
14	Happy	High
15	Unhappy	Low
16	Unhappy	Low
17	Happy	High
18	Happy	High
19	Unhappy	Low
20	Unhappy	Low

Climate change is destroying the polar bear's habitat and may well drive it to extinction. Human activity is not just correlated with global warming; it is one of the causes.

STUART YATES/ALAMY

FIGURE 2.2 **Causes of Correlation.** *If X (watching televised violence) and Y (aggressiveness) are correlated, then there are exactly three possible explanations: X causes Y, Y causes X, or Z (some other factor, such as lack of adult supervision) causes both Y and X, neither of which causes the other.*



COURTESY OF JIM SHAW

Although people have smoked tobacco for centuries, only recently has the causal relationship between cigarette smoke and lung disease been detected. By the way, how many physicians said the opposite? And "less irritating" than what?

and to watch televised violence (*X*) (**FIGURE 2.2**). For example, lack of adult supervision (*Z*) may allow children to get away with bullying others and to get away with watching television shows that adults would normally not allow. If this were true, then watching televised violence (*X*) and behaving aggressively (*Y*) may not be causally related to each other at all and may instead be the independent effects of a lack of adult supervision (*Z*), just as sneezing and coughing may be independent effects of viral infection, height and weight may be independent effects of nutrition, and so on. In other words, the relation between aggressiveness and televised violence may be a case of **third-variable correlation**, which means that *two variables are correlated only because each is causally related to a third variable*. How can we tell whether this is the case?

Matched Samples and Matched Pairs

The most straightforward way to determine whether a third variable such as lack of adult supervision (*Z*) causes children to watch televised violence (*X*) and to behave aggressively (*Y*) is to eliminate differences in adult supervision (*Z*) among a sample of children and see if the correlation between televised violence (*X*) and aggressiveness (*Y*) remains. For example, you could observe children using the **matched samples technique**, which is *a technique whereby the participants in two samples are identical in terms of a third variable*. For instance, we could measure only children who are supervised by an adult exactly 87% of the time, thus ensuring that every child who watched a lot of televised violence had exactly the same amount of adult supervision as every child who did not watch a lot of televised violence. Alternatively, you could observe children using the **matched pairs technique**, which is *a technique whereby each participant in a sample is identical to one other participant in that sample in terms of a third variable*. For instance, you could measure children who experience different amounts of adult supervision, but you could make sure that for every child you measure who watches a lot of televised violence and is supervised 24% of the time, you also observe a child who watches little televised violence and is supervised 24% of the time, thus ensuring that the children who do and do not watch a lot of televised violence have the same amount of adult supervision *on average*. Regardless of which technique you used, you would know that the children who do and don't watch televised violence have equal amounts of adult supervision on average; as such, if those who watch a lot of televised violence were more aggressive on average than those who didn't, then lack of adult supervision could not possibly be the cause.

third-variable correlation The fact that two variables may be correlated only because they are both caused by a third variable.

matched samples An observational technique that involves matching the average of the participants in the experimental and control groups in order to eliminate the possibility that a third variable (and not the independent variable) caused changes in the dependent variable.

matched pairs An observational technique that involves matching each participant in the experimental group with a specific participant in the control group in order to eliminate the possibility that a third variable (and not the independent variable) caused changes in the dependent variable.

third-variable problem The fact that the causal relationship between two variables cannot be inferred from the correlation between them because of the ever-present possibility of third-variable correlation.

ANDREANNA SEYMORE/STONE/GETTY

A woman's age is correlated with the number of children she has borne, but age does not cause women to become pregnant, and pregnancy does not cause women to age.

[HOT SCIENCE]

Establishing Causality in the Brain

Sometimes the best way to learn about something is to see what happens when it breaks, and the human brain is no exception. Scientists have studied the effects of brain damage for centuries, and those studies reveal a lot about how the brain normally works so well. But the problem with studying brain-damaged patients, of course, is the problem with studying any naturally occurring variable: Brain damage may be related to particular patterns of behavior, but that relationship may or may not be causal. Experimentation is the premiere method for establishing causal relationships between variables, but scientists cannot ethically cause brain damage in human beings, and thus they have not been able to establish causal relationships between particular kinds of brain damage and particular patterns of behavior.

Until now. Scientists have recently discovered a way to mimic brain damage with a benign technique called *transcranial magnetic stimulation* (or TMS) (Barker, Jalinous, & Freeston, 1985; Hallett, 2000). If you've ever held a magnet under a piece of paper and used it to drag a pin across the paper's surface, you know that magnetic fields can pass through material such as paper and wood. They can pass through bone, too. TMS delivers a magnetic pulse that passes through the skull and deactivates neurons in the cerebral cortex for a short period. Researchers can direct TMS pulses to particular brain regions—essentially turning them "off"—and then measure temporary changes in the way a person moves, sees, thinks, remembers, speaks, or feels. By manipulating the state of the brain, scientists can perform experiments that establish causal relationships. For example, scientists have recently discovered that magnetic stimulation of the visual cortex temporarily impairs a person's ability to detect the motion of an object, without impairing the person's ability to recognize that object (Beckers & Zeki, 1995). This intriguing discovery suggests that motion perception and object recognition are accomplished by different parts of the brain, but moreover, it establishes that the activity of these brain regions *causes* motion perception and object recognition.

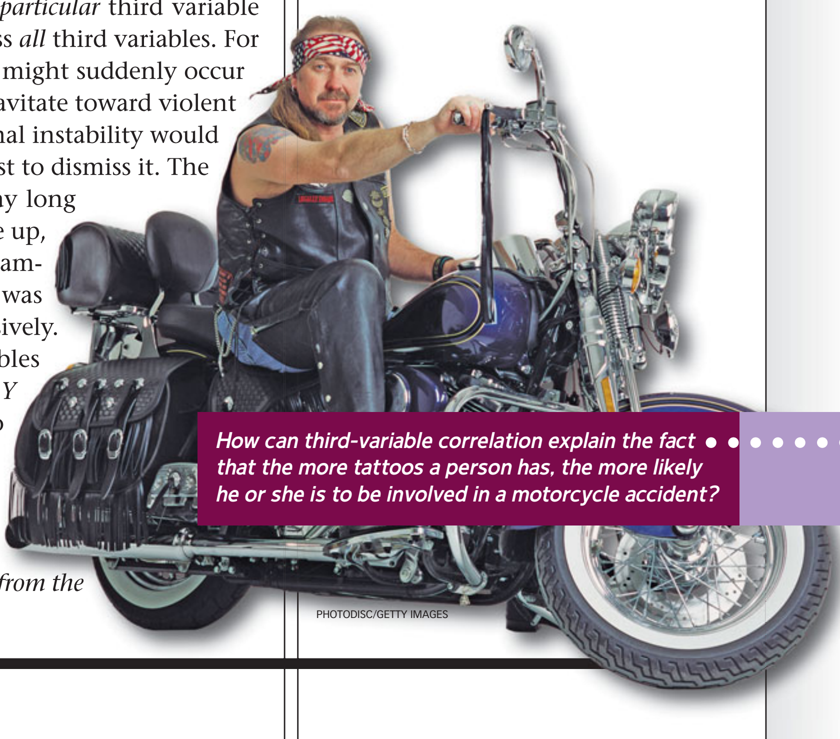

ARCELLO MASSIMINI/UNIVERSITY OF WISCONSIN-MADISON

Transcranial magnetic stimulation activates and deactivates regions of the brain with a magnetic pulse, temporarily mimicking brain damage.

For the first time in human history, the causal relationships between particular brain regions and particular behaviors have been unequivocally established. Rather than relying on observational studies of brain-damaged patients or the snapshots provided by MRI or PET scans, researchers can now manipulate brain activity and measure its effects. Studies suggest that TMS has no harmful side effects (Pascual-Leone et al., 1993), and this new tool promises to revolutionize the study of how our brains create our thoughts, feelings, and actions.

Although both the matched samples and matched pairs techniques can be useful, neither would allow you to dismiss the possibility of third-variable correlation entirely. Why? Because even if you used these techniques to dismiss a *particular* third variable (e.g., lack of adult supervision), you would not be able to dismiss *all* third variables. For example, as soon as you finished making these observations, it might suddenly occur to you that emotional instability (*Z*) might cause children to gravitate toward violent television programs (*X*) and to behave aggressively (*Y*). Emotional instability would be a new third variable, and you would have to design a new test to dismiss it. The problem is that you could dream up new third variables all day long without ever breaking a sweat, and every time you dreamed one up, you would have to rush out and do a new test using matched samples or matched pairs to determine whether *this* third variable was the cause of watching televised violence and of behaving aggressively.

The fact is that there are an infinite number of third variables out there and thus an infinite number of reasons why *X* and *Y* might be correlated. Because most of us don't have the time to perform an infinite number of studies with matched samples or matched pairs, we can never be sure that the correlation we observe between *X* and *Y* is evidence of a causal relationship between them. The **third-variable problem** refers to the fact that *a causal relationship between two variables cannot be inferred from the*



How can third-variable correlation explain the fact that the more tattoos a person has, the more likely he or she is to be involved in a motorcycle accident?

PHOTODISC/GETTY IMAGES

naturally occurring correlation between them because of the ever-present possibility of third-variable correlation. In other words, if we care about causality, then naturally occuring correlations can never tell us what we really want to know.

Experimentation

The third-variable problem prevents us from using naturally occuring correlations to learn about causal relationships, and so we have to find another method that will. Let's start by considering once again the source of all our troubles. We cannot conclude that watching televised violence causes children to behave aggressively because there is some chance that both behaviors are caused by a third variable, such as lack of adult supervision or emotional instability, and there are so many third variables in the world that we could never do enough tests to dismiss them all. Another way of saying this is that children who do watch and don't watch televised violence differ in countless ways, and any one of these countless differences could be the real cause of their different levels of aggressiveness and their different levels of violence watching. Of course, if we could somehow eliminate *all* of these countless differences at once—somehow find a sample of children who were perfect clones of each other, with identical amounts of adult supervision, identical amounts of emotional stability, identical histories, identical physiologies, identical neighborhoods, siblings, toys, schools, teeth, dreams, and so on—then we *could* conclude that televised violence and aggressiveness have a causal relationship. If we could only find a sample of children, some of whom watch televised violence and some of whom don't, but all of whom are identical in terms of *every possible* third variable, then we would know that watching televised violence and behaving aggressively are not just correlated but causally related.

CHILD: IVAN JOSIFOVIC/ISTOCKPHOTO; TV: IMAGERY MAJESTIC/DREAMSTIME.COM

• *Children's aggressiveness is correlated with the amount of violence they see on TV, but that doesn't mean that one of these things causes the other.*

Finding a sample of clones is not very likely, and so scientists have developed a technique that can eliminate all the countless differences between people in a sample. It is called an experiment. An **experiment** is *a technique for establishing the causal relationship between variables*. The most important thing to know about experiments is that you already know the most important thing about experiments because you've been doing them all your life. Imagine, for instance, what you would do if you were surfing the web on a laptop that used a wireless connection when all of a sudden the connection stopped working. You might suspect that another device, such as your roommate's new cordless phone, was interfering with your connection. Your first step would be to observe and measure carefully, noting whether you had a connection when your roommate was and was not using his cordless phone. But even if you observed a correlation between the failure to connect and your roommate's phone usage, the third-variable problem would prevent you from drawing a causal conclusion. For example, maybe your roommate is afraid of loud noises and calls his mommy for comfort whenever there is an electrical storm. And maybe the storm somehow interrupts your wireless connection. In other words, it is possible that a storm (*Z*) is the cause of both your roommate's phone calls (*X*) and your laptop's failure to connect to the internet (*Y*).

● In what ways do we perform experiments in everyday life?

How could you solve the third-variable problem? Rather than *observing* the correlation between telephone usage and connection failure, you could try to *create* a correlation by intentionally switching your roommate's phone on and off a few times and observing changes in your laptop's connection. If you noticed that "telephone on" and "connection failed" occurred together more often than not, then you would conclude that your roommate's telephone was the *cause* of your failed connection, and you would put the phone in the trash compactor and deny it when your roommate asked.

The technique you might intuitively use to solve your connection problem is the same technique that psychologists use to solve scientific problems. Consider again the

experiment A technique for establishing the causal relationship between variables.

independent variable The variable that is manipulated in an experiment.

experimental group One of the two groups of participants created by the manipulation of an independent variable in an experiment; the experimental group is exposed to the stimulus being studied and the *control group* is not.

control group One of the two groups of participants created by the manipulation of an independent variable in an experiment that is not exposed to the stimulus being studied.

dependent variable The variable that is measured in a study.

internal validity The characteristic of an experiment that allows one to draw accurate inferences about the causal relationship between an independent and dependent variable.

correlation between aggressiveness and televised violence. How can you determine why these variables are correlated? Well, rather than *measuring* how much televised violence a child watches (as you did when you used the matched pairs or matched sample techniques), you could *manipulate* how much televised violence a child watches. For example, you could find a sample of children, expose half of them to 2 hours of televised violence every day for a month, and make sure that the other half saw no televised violence at all (see FIGURE 2.3). At the end of a month, you could measure the aggressiveness of the children in the two groups. In essence, you would be computing the correlation between a variable that you measured (aggressiveness) and a variable that you manipulated (televised violence), and in so doing, you would have solved the third-variable problem. Because you *manipulated* rather than *measured* how much televised violence a child saw, you would never have to wonder whether a third variable (such as lack of adult supervision) might have caused it. Why? Because you already *know* what caused the child to watch or not watch televised violence. You did!

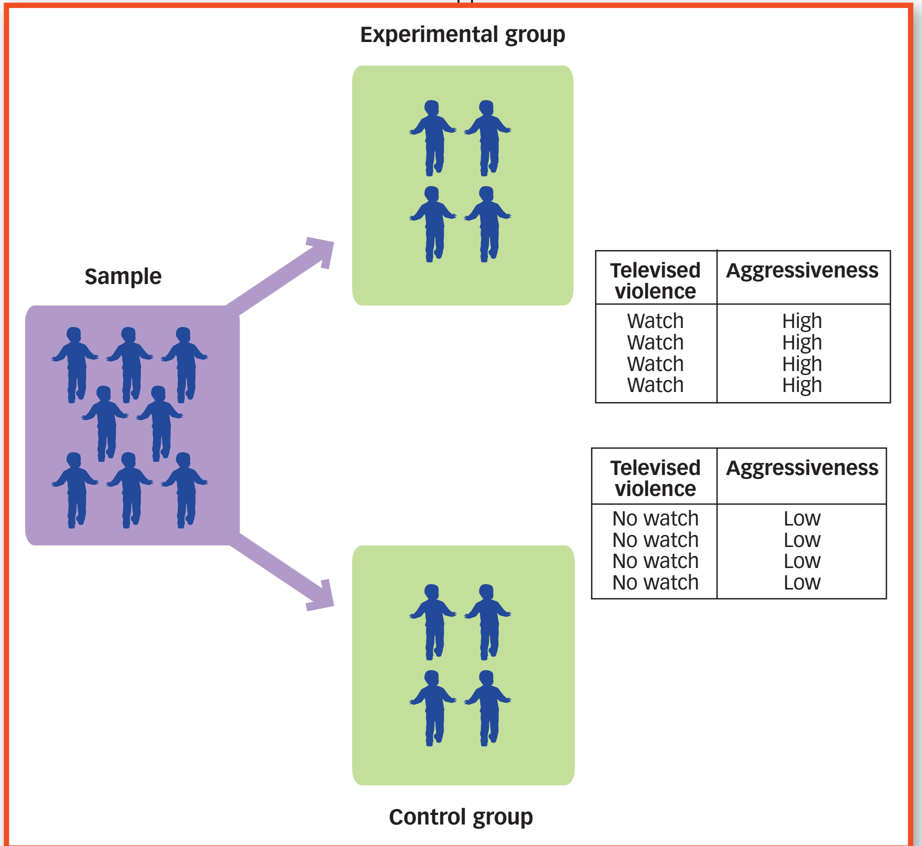


FIGURE 2.3 **Manipulation.** *The independent variable is televised violence and the dependent variable is aggressiveness. Manipulation of the independent variable results in an experimental group and a control group. When we compare the behavior of participants in these two groups, we are actually computing the correlation between the independent variable and the dependent variable.*

Experiments always involves manipulation. We call *the variable that is manipulated* the **independent variable** because it is under our control, and thus it is "independent" of what the participant says or does. When we manipulate an independent variable (e.g., watching televised violence), we create at least two groups of participants: an **experimental group**, which is *the group of people who are treated in a particular way,* such as being exposed to two hours of televised violence per day for a month, and a **control group**, which is *the group of people who are not treated in this particular way*. Then we measure another variable, and we call *the variable that is measured* the **dependent variable** because its value "depends" on what the participant says or does.

Drawing Conclusions

If you were to apply the techniques discussed so far, you could design an experiment that has **internal validity**, which is *the characteristic of an experiment that allows one to draw accurate inferences about the causal relationship between an independent and dependent variable*. When we say that an experiment is internally valid, we mean that everything *inside* the experiment is working exactly as it must in order for us to draw conclusions about causal relationships. Specifically, an experiment is internally valid when

- An independent variable has been effectively manipulated.
- A dependent variable has been measured in an unbiased way with a valid, powerful, and reliable measure.
- A correlation has been observed between the independent and the dependent variable.

If we do these things, then we may conclude that manipulated changes in the independent variable caused measured changes in the dependent variable. For example, our imaginary experiment on televised violence and aggressiveness would allow us conclude that televised violence (as we defined it) caused aggressiveness (as we defined it) in the

AP PHOTO/KEYSTONE, TIPRESS/SAMUEL GOLAY

STEWART FEREBEE/PHOTONICA/GETTY

Does piercing make a person more or less attractive? The answer, of course, depends entirely on how you operationally define piercing.

people whom we measured. Notice that the phrase "as we defined it" represents an important restriction on the kinds of conclusions we can draw from this experiment. Our experiment does not allow us to draw the sweeping conclusion that "televised violence causes aggressiveness" because as you saw when we discussed operational definitions, there are many different ways to define "televised violence" (should a shouting match between a Democratic and Republican strategist count as violence?) and many different ways to define "aggressiveness" (does cutting someone off midsentence count as aggressiveness?). Whether we observe a correlation between the violence we manipulated and the aggressiveness we measured will surely depend on how we defined them in the first place.

So what is the right way to define such variables? One obvious answer is that we should define them as they are typically defined in the real world. **External validity** is *a property of an experiment in which variables have been operationally defined in a normal, typical, or realistic way*. It seems fairly clear that *interrupting* is not the kind of aggressive behavior with which teachers and parents are normally concerned and that most instances of aggression among children lie somewhere between an insult and a chain saw massacre. If the goal of an experiment is to determine whether the kinds of programs children typically watch cause the kinds of aggression in which children typically engage, then external validity is essential.

Should variables be defined as they typically are in the real world?

But that isn't what many experiments are meant to do. Psychologists are rarely trying to learn about the real world by creating tiny replicas of it in their laboratories. Rather, they are usually trying to learn about the real world by using experiments to test theories and hypotheses (Mook, 1983). A **theory** is *a hypothetical account of how and why a phenomenon occurs*, and a **hypothesis** is *a testable prediction made by a theory*. For example, physicists have a theory stating that heat is the result of the rapid movement of molecules. This theory suggests a hypothesis—namely, that if the molecules that constitute an object are slowed, the object should become cooler. Now imagine that a physicist tested this hypothesis by performing an experiment in which a laser was used to slow the movement of the molecules in a rubber ball, whose temperature was then measured. Would we criticize this experiment by saying, "Sorry, but your experiment teaches us nothing about the real world because in the real world, no one actually uses lasers to slow the movement of the molecules in rubber balls"? Let's hope not. The physicist's theory (molecular motion causes heat) led to a hypothesis about what would happen in the laboratory (slowing the molecules in a rubber ball should cool it), and thus the events that the physicist manipulated and measured in the laboratory served to test the theory. Similarly, a good theory about the causal relationship between watching violence on television and behaving aggressively should lead to hypotheses about how children will behave after watching 2 minutes of Road Runner cartoons or all five *Nightmare on Elm Street* movies back to back. As such, even these this unrepresentative forms of television watching can serve to test the theory. In short, theories allow us to generate hypotheses about what *can* happen, or what *must* happen, or what *will* happen under particular circumstances, and experiments are typically meant to create these circumstances, test the hypotheses, and thereby provide evidence for or against the theories that generated them. Experiments are not meant to be miniature versions of everyday life, and thus external invalidity is not necessarily a problem.

"Hi. You've been randomly selected to participate in a sex survey upstairs in fifteen minutes."

Our imaginary experiment on televised violence and aggressiveness would allow us to conclude that televised violence (as we defined it) caused aggressiveness (as we defined it) in the people whom we measured. The phrase “the people whom we measured” represents another important restriction on the kinds of conclusions we can draw from this experiment. All experiments are done with a sample of participants who are drawn from a larger population. How can we know whether the conclusions we draw about the sample are also true of the larger population? The best way to do this is to use **random sampling**, which is *a technique for choosing participants that ensures that every member of a population has an equal chance of being included in the sample*. When we randomly sample participants from a population, we earn the right to *generalize*—that is, to conclude that what we observed in our sample would also have been observed if we had measured the entire population. You already have good intuitions about the importance of random sampling. For example, if you stopped at a farm stand to buy a bag of cherries and the farmer offered to let you taste a few that he had specially handpicked from the bag, you’d be reluctant to generalize from that nonrandom sample to the population of cherries in the bag. But if the farmer invited you to pull a few cherries from the bag without looking, you’d probably be willing to take those cherries as reasonably representative of the cherry population, and you’d be reasonably sure that your conclusions about these randomly sampled cherries would apply to the rest of the cherries in the bag.

Given the importance of random sampling, you may be surprised to learn that psychologists almost never do it. Indeed, virtually every participant in every psychology experiment you will ever read about was a volunteer, and most were college students who were significantly younger, smarter, healthier, wealthier, and whiter than the average earthling. Psychologists sample their participants the “wrong way” (by nonrandom sampling) because it is just about impossible to do it the “right way” (by random sampling). Even if there were an alphabetized list of all the world’s human inhabitants from which we could randomly choose our research participants, the likelihood that we could actually perform experiments on those whom we sampled would be depressingly slim. After all, how would we find the 72-year-old Bedouin woman whose family roams the desert so that we could measure the electrical activity in her brain while she watched cartoons? How would we convince the 3-week-old infant in New Delhi to complete a lengthy questionnaire about his political beliefs? Most psychology experiments are conducted by professors and graduate students at colleges and universities in the Western Hemisphere, and as much as they might like to randomly sample the population of the planet, the practical truth is that they are pretty much stuck studying the folks who volunteer for their studies.

● When can a sample teach us about a population?

So how can we learn *anything* from psychology experiments? Isn’t the failure to randomly sample a fatal flaw? No, it’s not. Although we can’t automatically generalize from nonrandom samples, there are three reasons why this is not a lethal problem for the science of psychology:

- Sometimes generality does not matter. One flying pig utterly disproves most people’s theories of porcine locomotion. Similarly, in psychology it often doesn’t matter if *everyone* does something as long as *someone* does it. If watching a violent television show for 1 hour caused a nonrandomly selected group of children to start shoving in the lunch line, then this fact would utterly disprove every theory that claimed that televised violence cannot cause aggression—and it might even provide important clues about when aggression will and won’t occur. An experimental result can be illuminating even when its generality is severely limited.
- Sometimes generality can be determined. When the generality of an experimental result *is* important, psychologists often perform a new experiment that uses the same procedures on a different sample.

external validity A characteristic of an experiment in which the independent and dependent variables are operationally defined in a normal, typical, or realistic way.

theory A hypothetical account of how and why a phenomenon occurs, usually in the form of a statement about the causal relationship between two or more properties. Theories lead to *hypotheses*.

hypothesis A specific and testable prediction that is usually derived from a *theory*.

random sampling A technique for choosing participants that ensures that every member of a population has an equal chance of being included in the sample.


GETTY IMAGES

College students are the traditional "guinea pigs" of psychological research.

For example, if we were to measure how some American children behaved after watching televised violence for 2 hours, we could then replicate the experiment with Japanese children, or with teenagers, or with adults. In essence, we could treat the attributes of our sample, such as culture and age, as independent variables and do experiments to determine whether these attributes influenced our dependent variable. If the results of our study were replicated in numerous nonrandom samples, we could be more confident (though never completely confident) that the results would generalize to the population at large.

- Sometimes generality can be assumed. Instead of asking, "Is there a compelling reason to generalize from a nonrandom sample?" we might just as easily ask, "Is there a compelling reason not to?" For example, few of us would be willing to take an experimental drug that could potentially make us smarter and happier if a nonrandom sample of seven participants took the drug and died a slow, painful death. Indeed, we would probably refuse the drug even if the seven subjects were mice. Although the study used a nonrandom sample of participants who are different from us in many ways, we are willing to generalize from their experience to ours because we know that even mice share enough of our basic biology to make it a good bet that what harms them can harm us, too. By this same reasoning, if a psychology experiment demonstrated that some American children behaved violently after watching televised violence for 1 hour, we might ask whether there is a compelling reason to suspect that Ecuadorian college students or middle-aged Australians would behave any differently. If we had a reason to suspect they would, then the experimental method would provide a way for us to investigate that possibility.

summary quiz [2.2]

5. Your friend tells you that she has just heard that there is a positive correlation between pizza consumption and children's intelligence. If this were in fact true, it would mean that

a. the more intelligent children are, the more pizza they eat.
b. pizza ingredients enhance brain development.
c. intelligent children realize that pizza is a healthful food.
d. parents of intelligent children encourage them to eat lots of pizza.

6. Marie Rodriquez divides her seventh-grade gifted class in half during study hour. Half watch a video encouraging volunteer activities, while the other half watch an MTV video. She then records how aggressively students behave at recess later that day. What is the independent variable in this study?

a. students' aggressive behavior
b. recess
c. giftedness of the students
d. type of video

7. Dr. Shondra Jones administers to female and male students at both northern universities and southern universities a questionnaire that measures attitudes towards women's rights. The dependent variable in this study is

a. gender of the students.
b. geographic location of the students.
c. attitudes toward women's rights.
d. the majors of the students.

8. The characteristic of an experiment that allows one to draw accurate inferences about the causal relationship between an independent and dependent variable is called

a. external validity.
b. internal validity.
c. third-variable correlation.
d. matched sample technique.

The Ethics of Science: Saying Please and Thank You

Somewhere along the way, someone probably told you that it isn't nice to treat people like objects. And yet, it may seem that psychologists do just that—creating situations that cause people to feel fearful or sad, to do things that are embarrassing or immoral, and to learn things about themselves that they might not really want to know. Why do psychologists treat people so shabbily? In fact, psychologists go to great lengths to ensure the safety and well-being of their research participants, and they are bound by a code of ethics that is as detailed and demanding as the professional codes that bind physicians, lawyers, and members of the clergy. This code of ethics was formalized by the American Psychological Association in 1958 and offers a number of rules that govern all research conducted with human beings. Here are a few of the most important ones:

- *Informed consent:* Participants may not take part in a psychological study unless they have given **informed consent**, which is a *written agreement to participate in a study made by an adult who has been informed of all the risks that participation may entail*. This doesn't mean that the person must know everything about the study (the hypothesis), but it does mean that the person must know about anything that might potentially be harmful, painful, embarrassing, or unpleasant. If people cannot give informed consent (perhaps because they are minors or are mentally incapable), then informed consent must be obtained from their legal guardians.
- *Freedom from coercion:* Psychologists may not coerce participation. Coercion not only means physical and psychological coercion but monetary coercion as well. It is unethical to offer people large amounts of money to persuade them to do something that they might otherwise decline to do. College students may be invited to participate in studies as part of their training in psychology, but they are ordinarily offered the option of learning the same things by other means.
- *Protection from harm:* Psychologists must take every possible precaution to protect their research participants from physical or psychological harm. If there are two equally effective ways to study something, the psychologist must use the safer method. If no safe method is available, the psychologist may not perform the study.
- *Risk-benefit analysis:* Although participants may be asked to accept small risks, such as a minor shock or a small embarrassment, they may not even be *asked* to accept large risks, such as severe pain or psychological trauma, or risks that are greater than those they would ordinarily take in their everyday lives. Furthermore, even when participants are asked to take small risks, the psychologist must first demonstrate that these risks are outweighed by the social benefits of the new knowledge that might be gained from the study.
- *Debriefing:* Although psychologists need not divulge everything about a study before a person participates, they must divulge it after the person participates. If a participant is deceived in any way before or during a study, the psychologist must provide a **debriefing**, which is *a verbal description of the true nature and purpose of a study.* If the participant was changed in any way (e.g., made to feel sad), the psychologist must attempt to undo that change (e.g., ask the person to do a task that will make them happy) and restore the participant to the state he or she was in before the study.

These rules require that psychologists show extraordinary concern for their participants' welfare, but how are they enforced? Almost all psychology studies are done by psychologists who work at colleges and universities. These institutions have institutional review boards (IRBs) that are composed of instructors and researchers, university

informed consent A written agreement to participate in a study made by a person who has been informed of all the risks that participation may entail.

debriefing A verbal description of the true nature and purpose of a study that psychologists provide to people after they have participated in the study.

"I don't usually volunteer for experiments, but I'm kind of a puzzle freak."

ONLY HUMAN

THE WELL-BEING OF PARTICIPANTS ALWAYS COMES FIRST! In 1997 in Mill Valley, California, 10th-grade student Ari Hoffman won first place in the Marin County science fair for doing a study that found that exposure to radiation decreased the offspring of fruitflies. However, he was quickly disqualified for cruelty when it was learned that about 35 of his 200 flies died during the 3-month experiment. Hoffman was disappointed because he had used extraordinary efforts to keep the flies alive, for example, by maintaining a tropical temperature for his flies during the entire experiment.

staff, and laypeople from the community (e.g., business leaders or members of the clergy). A psychologist may conduct a study only after the IRB has reviewed and approved it. As you can imagine, the code of ethics and the procedure for approval are so strict that many studies simply cannot be performed anywhere, by anyone, at any time. For example, psychologists have long wondered how growing up without exposure to language affects a person's subsequent ability to speak and think, but they cannot ethically manipulate such a variable in an experiment. As such, they must be content to study the natural correlations between variables such as language exposure and speaking ability, and they must forever forgo the possibility of firmly establishing causal relationships between these variables. There are many questions that psychologists will never be able to answer definitively because doing so would require unethical experimentation. This is an unavoidable consequence of studying creatures who have fundamental human rights.

● Is it ever justifiable to harm a human or nonhuman research participant?

Of course, not all research participants have human rights because not all research participants are human. Some are chimpanzees, rats, pigeons, or other nonhuman animals. How does the ethical code of the psychologist apply to nonhuman participants? The question of animal rights is one of the most hotly debated issues of our time, and people on opposite sides of the debate rarely have much good to say about each other. And yet, consider three points on which every reasonable person would agree:

- A very small percentage of psychological experiments are performed on nonhuman animals, and a very small percentage of these experiments cause discomfort or death.
- Nonhuman animals deserve good care, should never be subjected to more discomfort than is absolutely necessary, and should be protected by federal and institutional guidelines.
- Some experiments on nonhuman animals have had tremendous benefits for human beings, and many have not.

None of these points is in dispute among thoughtful advocates of different positions, so what exactly is the controversy? The controversy lies in the answer to a single question: Is it morally acceptable to force nonhuman animals to pay certain costs so that human animals can reap uncertain benefits? Although compelling arguments may be made on both sides of this moral dilemma, it is clearly just that—a *moral* dilemma and not a scientific controversy that one can hope to answer with evidence and facts. Anyone who has ever loved a pet can empathize with the plight of the nonhuman animal that is being forced to participate in an experiment, feel pain, or even die when it would clearly prefer not to. Anyone who has ever loved a person with a debilitating illness can understand the desire of researchers to develop drugs and medical procedures by doing to nonhuman animals the same things that farmers and animal trainers do every day. Do animals have rights, and if so, do they ever outweigh the rights of people? This is a difficult question with which individuals and societies are currently wrestling. For now, at least, there is no easy answer.

PAUL MCERLANE/REUTERS/CORBIS

● *Some people consider it unethical to use animals for clothing or research. Others see an important distinction between these two purposes.*

summary quiz [2.3]

9. A written agreement to participate in a study made by an adult who has been informed of all the risks that participation may entail is known as

a. a memorandum of understanding.
b. informed consent.
c. a signature of authorization.
d. debriefing.

10. Participants in a psychology study
   a. can be asked to accept risks such as severe pain or psychological trauma.
   b. can be asked to accept risks that are greater than those they would ordinarily take in their everyday lives.
   c. can be asked to accept small risks, whether or not these are outweighed by the social benefits of gaining new knowledge.
   d. can be asked to accept small risks, but only if these risks are outweighed by the social benefits of gaining new knowledge.

11. A verbal description of the true nature and purpose of a study provided to the participant after the study is done is called
   a. informed consent.
   b. risk-benefit analysis.
   c. debriefing.
   d. harm-protection follow-up.

12. According to the textbook, reasonable people would agree on which of the following points regarding research performed on nonhuman animals?
   a. Some experiments on nonhuman animals have had tremendous benefits for humans.
   b. Nonhuman animals should never be subjected to discomfort or death even if the results would greatly benefit humans.
   c. Although human research participants are protected by federal and institutional guidelines, nonhuman animals do not need these protections.
   d. It can be demonstrated scientifically that the rights of humans always outweigh the rights of nonhuman animals.

WhereDoYouStand?

The Morality of Immoral Experiments

Is it wrong to benefit from someone else's wrongdoing? Although this may seem like an abstract question for moral philosophers, it is a very real question that scientists must ask when they consider the results of unethical experiments. During World War II, Nazi doctors conducted barbaric medical studies on prisoners in concentration camps. They placed prisoners in decompression chambers and then dissected their living brains, in order to determine how similar decompression at high altitude might affect pilots. They irradiated and chemically mutilated the reproductive organs of men and women in order to find inexpensive methods for the mass sterilization of "racially inferior" people. They infected prisoners with streptococcus and tetanus in order to devise treatments for soldiers who had been exposed to these bacteria. And in one of the most horrible experiments, prisoners were immersed in tanks of ice water so that the doctors could discover how long pilots would survive if they bailed out over the North Sea. The prisoners were frozen, thawed, and frozen again until they died. During these experiments, the doctors carefully recorded the prisoners' physiological responses.

These experiments were crimes, but the records of these experiments remain, and in some cases they provide valuable information that could never be obtained ethically. For example, because researchers cannot perform controlled studies that would expose volunteers to dangerously cold temperatures, there is still controversy among doctors about the best treatment for hypothermia. In 1988, Robert Pozos, a physiologist who had spent a lifetime studying hypothermia, came across an unpublished report written in 1945 titled "The Treatment of Shock from Prolonged Exposure to Cold, Especially in Water." The report described the results of the horrible freezing experiments performed on prisoners at the Dachau concentration camp, and it suggested that contrary to the conventional medical wisdom, rapid rewarming (rather than slow rewarming) might be the best way to treat hypothermia.

Should the Nazi medical studies have been published so that modern doctors might more effectively treat hypothermia? Many scientists and ethicists thought they should. "The prevention of a death outweighs the protection of a memory. The victims' dignity was irrevocably lost in vats of freezing liquid forty years ago. Nothing can change that," argued bioethicist Arthur Caplan. Others disagreed. "I don't see how any credence can be given to the work of unethical investigators," wrote Arnold Relman, editor of the *New England Journal of Medicine*. "It goes to legitimizing the evil done," added Abraham Foxman, national director of the Anti-Defamation League (Siegel, 1988). The debate about this issue rages on (Caplan, 1992). If we use data that were obtained unethically, are we rewarding those who collected it and legitimizing their actions? Or can we condemn such investigations but still learn from them? Where do you stand?

CHAPTER REVIEW

Summary

Empiricism: How to Know Things

- An empiricist believes observation is key to accurate knowledge.
- A method is a set of rules and techniques for observation necessary to avoid mistakes that simple observation can produce.

The Science of Observation: Saying What

- Measurement is a scientific means of observation that involves defining an abstract property in terms of some concrete condition, called an operational definition, and then constructing a device, or a measure, that can detect the conditions that the operational definition specifies.
- Psychologists sometimes use the case method to study single, exceptional individuals, but more often they use samples of many people drawn from a population.
- When people know they are being observed, they may behave as they think they should; psychologists try to reduce or eliminate such demand characteristics by observing participants in their natural habitat or by hiding their expectations from people.
- In double-blind observations, the experiment's purpose is hidden from the experimenter and the participants, ensuring that observers neither see what they want to see nor cause participants to behave as the observers expect them to behave.

The Science of Explanation: Saying Why

- To determine whether two variables are causally related, we must first determine whether they are related at all.
- Even when we find a correlation between two variables, we can't conclude that they are causally related, because an infinite number of "third variables" might be causing them both.
- Experiments solve the third-variable problem by manipulating an independent variable, assigning participants to the experimental and control groups that this manipulation creates, and measuring a dependent variable which is then compared across groups.
- An internally valid experiment establishes a causal relationship between variables as they were operationally defined and among the participants whom they included. When an experiment is externally valid—that is, when the variables mimic the real world and participants are randomly sampled—we may generalize from its results. Internal validity is essential; external validity is not.

The Ethics of Science: Saying Please and Thank You

- Psychologists have the responsibility of making sure that human research participants give their informed and voluntary consent to participate in studies, and that these studies pose minimal or no risk.
- Similar principles guide the human treatment of nonhuman research subjects.
- Enforcement of these principles by federal, institutional, and professional governing agencies ensures that the research process is a meaningful one that can lead to significant increases in knowledge.

Key Terms

empiricism (p. 34)
method (p. 34)
operational definition (p. 35)
measure (p. 35)
electromyograph (EMG) (p. 35)
validity (p. 36)
reliability (p. 36)
power (p. 36)
case method (p. 37)
population (p. 37)
sample (p. 37)
demand characteristics (p. 37)
naturalistic observation (p. 37)
double-blind observation (p. 39)
correlated (p. 41)
variable (p. 41)
third-variable correlation (p. 42)
matched samples (p. 42)
matched pairs (p. 42)
third-variable problem (p. 43)
experiment (p. 44)
independent variable (p. 45)
experimental group (p. 45)
control group (p. 45)
dependent variable (p. 45)
internal validity (p. 45)
external validity (p. 46)
theory (p. 46)
hypothesis (p. 46)
random sampling (p. 47)
informed consent (p. 49)
debriefing (p. 49)

Critical Thinking Questions

1. Among the ancient Greeks, the dogmatists were healers who tried to understand the body by developing theories about its function; in contrast, the empiricists tried to understand the body by observing sick people.

   Today, our modern word *dogmatic* is used to describe someone who authoritatively states his opinions as if they were facts, while an unqualified physician who trusts his own experience without regard to established theory or standard practice can be called a quack, a charlatan, or an "empiric."

   How do you think these modern definitions grew out of the ancient ones?

2. Demand characteristics are those aspects of a setting or experiment that cause people to behave as they think an observer wants or expects them to behave. Experimental results can also be skewed because the experimenter's expectations can influence observations.

   Suppose you are a medical researcher conducting a study to see if a new trial drug called Relievia is more or less effective than aspirin at relieving pain. On the one hand, although you plan to run the study fairly, you hope that Relievia will work, because you've invested so much time in developing this drug. On the other hand, the study participants may feel pressured to report that the fancy-sounding, expensive new drug works better than plain old aspirin. Both these demand characteristics and your own preconceptions could influence the results.

   How would you design your experiment to minimize the effects of these preconceptions on your study?

3. A fundamental idea in psychology is that correlation (two things that tend to occur together) does not necessarily imply causation (one thing causing the other). This is an idea that confuses many people (including some psychologists).

   For example, many newspaper articles have now noted that people who live in houses located under high-voltage power lines have a heightened risk of developing certain kinds of cancer. In other words, living under power lines and cancer risk may be statistically correlated.

   Does this mean that long-term exposure to power lines causes cancer? If not, what else might explain the correlation?

Answers to Summary Quizzes

Summary Quiz 2.1
1. b; 2. a; 3. c; 4. b

Summary Quiz 2.2
5. a; 6. d; 7. c; 8. b

Summary Quiz 2.3
9. b; 10. d; 11. c; 12. a

Need more help? Additional resources are located at the book's free companion Web site at:
www.worthpublishers.com/schacterbrief1e