
FABRICE COFFRINI/AFP via Getty Images

2 Research Methods

Core Questions

2.1 How do social psychologists design studies?

2.2 How do experiments work in social psychology?

2.3 How can I recognize trustworthy research?

Learning Objectives

2.1 Describe the scientific method, methods of descriptive data collection, and correlations.

2.2 Understand the strengths and weaknesses of experiments and how their results are analyzed.

2.3 Explain why reliability, validity, random sampling, ethics, and open science signal good research.

Like all sciences, social psychology usually moves like snail: steady but slow. It is slow, in part, because what social psychologists study is usually often invisible—and therefore difficult to measure. For example, prejudice, persuasion, altruism, and romantic love are all scientific **constructs**, theoretical ideas that cannot be directly observed. Although the scientific process is slow, social psychology is growing fast. It is growing fast because so many students are attracted to Kurt Lewin's vision of an applied science.

Perhaps social psychology's popularity explains why so many passengers were carrying long plastic or cardboard tubes on a recent plane ride. The plane was full of people presenting at a conference sponsored by the Society for Personality and Social Psychology (SPSP), which happens at the end of every winter. The tubes contained rolled-up posters summarizing the most cutting-edge research in the field. This chapter describes how all those professional established scientists, graduate students, and even a few undergraduates created those studies—and it invites you to join us.

HOW DO SOCIAL PSYCHOLOGISTS DESIGN STUDIES?

>> LO 2.1: Describe the scientific method, methods of descriptive data collection, and correlations.

The working world is full of designers.

We have fashion designers, graphic designers, architectural designers, cookware designers, landscape designers, and game designers. To become a clear-thinking social psychologist, you must become a research designer. Although there are guidelines, designing research is also an art that you can develop with experience—just like a fashion or a landscape designer.

The purpose of a research project often determines its design. The purpose of **basic science** is to increase understanding, create testable theories, and predict social behavior. Basic research strives to understand a given phenomenon more. The purpose of **applied science** is to translate those theories into applied problem solving or social action. Applied research is used to make the world better in some tangible way, like reducing aggression or increasing self-esteem.

Both are important. Basic research is important because, as social psychology's pioneer Kurt Lewin famously recognized, "There is nothing so practical as a good theory" (see Bedeian, 2016; Lewin, 1951, p. 169). Applied science is where theory wrestles with reality—but with the understanding that reality always wins. If a theory does not describe reality, then the theory has to change. Figure 2.1 displays the four phases of the scientific method used in both basic and applied research.

This section answers the core question "How do social psychologists design studies?" by

(1) describing the scientific method,

(2) comparing descriptive research designs, and

(3) explaining correlations.

Apply the Scientific Method

The whole **scientific method** is greater than the sum of its parts.

The parts of the scientific method include hypothesis generation, types of designs, procedural techniques, and measurement tools that are slightly different in every science. But the whole of the scientific method is an attitude based on healthy skepticism, the belief that evidence is the most trustworthy way to know about something. Healthy skepticism also implies a stubborn, Galileo-like refusal to believe something just because an authority says so. When opinion meets data, the data win.

Constructs: Theoretical ideas that cannot be directly observed, such as attitudes, personality, attraction, or how we think.

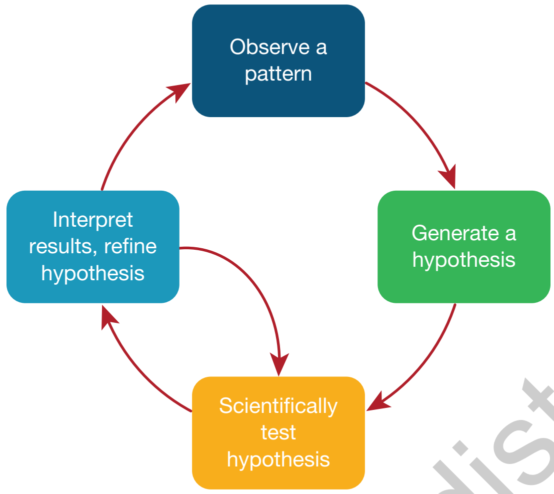
Basic science: Research that increases understanding and theory within a field like psychology.

Applied science: Research that translates theory into applied problem solving or social action.

Scientific method: A systematic way of creating knowledge by observing, forming a hypothesis, testing a hypothesis, and interpreting the results.

FIGURE 2.1

The scientific method includes these basic elements.



Most research stories cycle through the same four phases. This pattern of research, shown in Figure 2.1, is repeated with endless variations. You will grow accustomed to this rhythm: observation, hypothesis, testing, and interpretation of the results.

- *Phase 1: Observe a pattern of behavior.* Imagine that you are in a coffee shop quietly observing other customers. You notice that men frequently interrupt people during conversations—and that seems to be especially true when their conversation partner is a woman. Welcome aboard; you started the scientific journey when your curiosity prompted you to ask, "Is this a pattern?" (By the way, this exact observation was tested in coffee shops and drugstores back in 1975 by Zimmerman and West.) Phase 1 of the scientific method begins when we observe what we think is a reliable pattern of behavior. Of course, Phase 1 doesn't always start with literal observation in this way. It can also start with observation of studies already done on a topic, when you notice something that could be added to this line of research to enhance understanding.
- *Phase 2: Generate a hypothesis.* A formal **hypothesis** specifies what you believe will happen when you test your observations. It's an educated guess in the form of a statement. You might hypothesize that (1) men interrupt more than women do in general, and (2) men interrupt women more than they interrupt other men. Hypotheses are never stated as questions (such as, "Who interrupts more, men or women?"). They start as statements of what you think is going to happen, so that you can either reject the idea as wrong, gain support that you're right, or realize that the answer is more complicated than you originally thought.
- *Phase 3: Test the hypothesis.* You can do this in a wide variety of ways, which are described in the rest of this chapter. In this case, you'd probably use discreet observation techniques to watch people's behaviors in public places (like the coffee shop). But there are a lot of other options, and we'll cover those soon. And remember that a single study showing a pattern probably isn't enough to be really confident that your hypothesis was right; you need replication to build that confidence.
- *Phase 4: Interpret the results and refine your hypothesis.* If you found that your hypothesis was totally wrong, you might need to start all over at Phase 1. Or, you might realize that your hypothesis is true some of the time but not all of the time. In that case, you refine the hypothesis and test it again.

Hypothesis: A specific statement made by a researcher about the expected outcome of a study.

Imagine that we found support for the basic idea that, overall, men are indeed more likely than women to interrupt someone. Consider the following possible new, refined hypotheses:

- Women with more assertive personalities are more likely to interrupt others, compared to women with less assertive personalities.
- Men are less likely to interrupt women they find physically attractive, compared to women they don't find attractive.
- Men interrupt others more in friendly or informal settings, compared to formal settings such as at work.
- Men from cultures with more traditional gender roles are more likely to interrupt women than are men from more egalitarian cultures.

Any scientific data story can only unfold if our observations are as objective as possible, our hypotheses are specific, our tests are fair, and the results are properly interpreted. Any individual research study is a very small piece of a very big puzzle. Importantly, the methods and procedure chosen for any given study will have important implications for what conclusions can safely be drawn. Every completed study brings us a tiny bit closer to understanding the complicated world of social interaction.

Begin With a Descriptive Design

Many research projects begin with a descriptive design.

There are a lot of choices when it comes to choosing a way to gather data for a study, and we won't cover all of them here (it's not a research methods book). We'll start with three options that typically fall into the category of descriptive designs. **Descriptive designs** define, explain, and clarify patterns of people or events that happen without experimenter intervention.

The idea is that they would have happened anyway, and the researcher's study helps us describe those observations in a more detailed way. Imagine you are trying to understand how, when, and why some first-year college students drop out before their second year—and who is more likely to drop out compared to others. Here are three ways we could try to answer that question using descriptive designs.

Archival Data

Archival data are stored information that were originally gathered for a different purpose but can now be used to test hypotheses.

For example, most colleges and universities collect (and store) lots of information about their students. It includes application information (like high school GPA and hometowns), courses taken at that college, how long it took to graduate, what students majored in, and more. This is a lot of information, and patterns could be hypothesized and tested regarding whether certain types of students drop out more or less.

There are many sources of archival data. Newspapers may report quotations from people who witnessed important events like natural disasters—and those quotations could be analyzed. Census data might be used to track patterns such as how many people of a certain socioeconomic status are married, cohabiting, or live alone. Facebook profiles and posts are used by social psychologists to study how people reveal personal information about themselves. Police records contain data that can reveal patterns of reported domestic violence or other crimes.

Archival data are a rich source of information that might be waiting quietly for someone to analyze and reveal patterns of human behavior.

Naturalistic Observation

Another option is scientific surveillance.

Descriptive designs: Methods of gathering data that define, explain, and clarify patterns that happen without experimenter intervention.

Archival data: Stored information that was originally created for some other purpose that can later be used to test hypotheses, such as census or college records.

This descriptive design is usually called **naturalistic observation**, or watching people in their natural environments and recording their behaviors with a preset coding system, based on your hypothesis. By "natural," we don't mean in a cornfield or a forest but where the behavior normally occurs (such as a coffee shop). To find out what kinds of conversations college students have about dropping out, we might go to campus locations where we think we might overhear such conversations. This might be places like dorm lounges, academic success offices, free tutoring spots, or maybe even campus bars.

You may already notice that systems of data collection can sometimes overlap. If you decided to observe people's behaviors over several hours of videos that had to be created as security surveillance, you'd be using both naturalistic observation and archival data. You might also be using your critical thinking skills as you start to wonder about two important questions regarding naturalistic observation.

First, is naturalistic observation ethical? Video surveillance may have issues regarding the taped people's privacy (see Bhatia et al., 2019). For example, video surveillance is being developed in order to track unvaccinated individuals who, unknown to themselves, may be carrying a highly contagious, airborne disease into an international airport and then onto other countries.

Even without videos, observing people and using them for your own research purposes may be unethical because they haven't given consent to be part of your study. For this reason, companies, colleges, and universities always run their study ideas through an objective ethics committee before they engage in any research endeavors (this is described more later in the chapter). That said, most laws support the idea that if people are out in public, they are giving up their right to privacy.

Naturalistic observation: Watching and recording people's behaviors where they would have happened anyway, but for research purposes.

Reactivity: When people change their behavior because they realize they're being observed.

Participant observation: A technique used during naturalistic observation where scientists covertly disguise themselves as people belonging in an environment.

A second question you might be thinking is, "If people realize they're being observed, wouldn't they change their behaviors?" If that thought occurred to you, then congratulations—that is a legitimate scientific concern. **Reactivity** occurs when people's behavior changes when they know that they're being watched. They might become more polite, or try to show off, or say nice things about their boss, or do anything else they think makes them look like better people. When that happens, your study will have problems because the behavior you're trying to observe is no longer natural and authentic.

One possible solution to this problem is **participant observation**, which is when researchers go undercover and pretend to be part of the natural environment so that no one notices them. They might pretend to be a substitute teacher in a school, a maintenance worker at a park, or a member of a cult. In this way, hopefully the patterns

Pictorial Press Ltd / Alamy Stock Photo

Allstar Picture Library / Alamy Stock Photo

In the movie *21 Jump Street*, two young police officers go undercover pretending to be high school students, so they can bust a new drug that's hitting the community. In *Imperium*, Daniel Radcliffe's character works for the FBI and infiltrates a White supremacist group, pretending to be racist. If any of them had been social psychologists doing research with this undercover technique, it would have been called participant observation.

they observe and record will reflect genuine social interactions. The more authentic the behavior being observed, the better it is for research.

Descriptive Surveys

Sometimes the simplest way to collect data works best: Just ask.

Research surveys ask people to honestly report their thoughts, emotions, and behaviors. Often surveys include **self-report scales**, which ask people to respond to several items on a range (such as from 1 = *strongly disagree* to 7 = *strongly agree*). The researcher later sums or averages the responses to items and assigns that participant a score on a given variable. College administrators might ask current students to complete satisfaction scales as a way of trying to predict who might drop out versus make it all the way to graduation.

Creating a good survey is not as easy as it looks. Too few questions mean you won't get the information you desire. Too many questions mean participants will get bored, quit, or just make up information to get it over with. But when done well, there are many advantages to the survey method. Surveys

(a) are relatively inexpensive,

(b) can reach hundreds of people relatively quickly,

(c) can assess personal information in ways that may not be possible through naturalistic observations, and

(d) attract participants from anywhere in the world (especially if the survey is online). That's good because it means we can get a wide diversity of participants.

However, self-report surveys have a big problem: **social desirability**. This occurs when participants provide inaccurate information—they fudge their answers—in order to impress or please the researcher or simply because they don't want to admit something. For example, would you tell the truth if asked whether you've ever mistreated a romantic partner, cheated on a test, stolen something, or used illegal drugs? Do not despair! The What's My Score? feature describes one way to circumvent this problem when collecting data through surveys.

Self-report scale: A survey where participants give information about themselves by responding to several items along the same theme.

Social desirability: The tendency for participants to provide dishonest responses so that others have positive impressions of them or because they don't want to admit something.

WHAT'S MY SCORE?

Measuring Social Desirability

Social desirability damages the quality of self-report scales when participants answer in a way that they think makes them look good. One creative way around this problem is to include a measure of how willing participants are to be honest. Most people *have* done many of the bad behaviors listed in this scale. So, if research participants don't admit to any of them, then they are probably changing their answers on other parts of the survey to look good.

Instructions: Listed below are several statements concerning personal attitudes and traits. Please read each item and decide whether the statement is true or false as it pertains to you personally.

Circle "T" for true statements and "F" for false statements.

T F 1. Before voting, I thoroughly investigate the qualifications of all the candidates.

T F 2. I never hesitate to go out of my way to help someone in trouble.

T F 3. I sometimes feel resentful when I don't get my way.

T F 4. I am always careful about my manner of dress.

T F 5. My table manners at home are as good as when I eat out in a restaurant.

(Continued)

(Continued)

T F 6. I like to gossip at times.

T F 7. I can remember playing sick to get out of something.

T F 8. There have been occasions when I took advantage of someone.

T F 9. I'm always willing to admit it when I make a mistake.

T F 10. There have been occasions when I felt like smashing things.

T F 11. I am always courteous, even to people who are disagreeable.

T F 12. At times I have really insisted on having things my own way.

Scoring: Give yourself 1 point if you said TRUE for Items 1, 2, 4, 5, 9, or 11. Then, give yourself 1 point if you said FALSE for Items 3, 6, 7, 8, 10, or 12. The more points you have, the more you are trying to manage your impression—in other words, you have a higher score on social desirability.

Source: Crowne and Marlowe (1960).

Understand Correlational Analyses

For many students (including your authors), the **correlational analysis** was a door-opening experience into the surprisingly pleasant world of statistics.

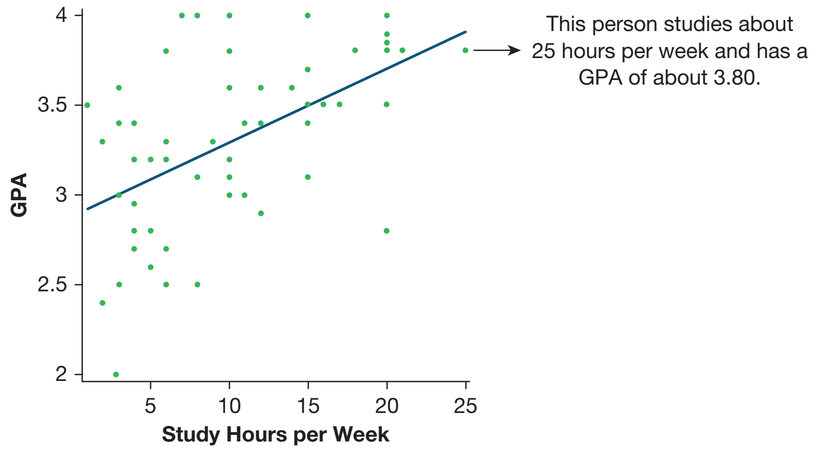
A correlation starts with two pieces of information from each participant. Each piece of information is a number that represents where that person falls on a range, or continuum, for two variables of interest. For example, you can ask people (1) how many hours they study each week and (2) their grade point average (GPA). The obvious hypothesis: More studying is associated with higher grades.

Correlational analysis: A statistic testing if two continuous variables are systematically associated with each other.

Figure 2.2 tells a beautifully nuanced yet easy-to-understand story. The pattern of dots flows upward and to the right, so the general theme is that more hours spent studying each week is associated with higher grades. And each individual dot is also

FIGURE 2.2

In this graph, each dot represents one person. For each person, study hours per week fall on the *x*-axis, and grade point average (GPA) falls on the *y*-axis. By looking at the general pattern, we can determine whether the two variables are correlated.



a personal story. People in the upper left get high grades without studying very much; people in the lower right get low grades but study a great deal.

Correlations in which both variables move in the same direction are called "positive" correlations. Here, as studying goes up, grades go up—or as studying goes down, grades go down. So studying and grades have a positive correlation. If the two variables move in opposite directions—as one goes up, the other goes down—it's called a "negative" correlation. You might hypothesize a negative correlation between partying and grades: More partying means lower grades.

Correlations will always be represented as a number that ranges from −1.00 to +1.00. You already know what the sign (positive or negative) means. The number represents how closely each dot on the graph follows the pattern, or how close the dots are to the line. If they are all *exactly* on the line, then the association between the two variables is perfect, and the number will be 1.00 (either positive or negative). As the number gets closer to zero (from either direction), the dots start to spread out. That means there's more variation, and the association isn't as strong.

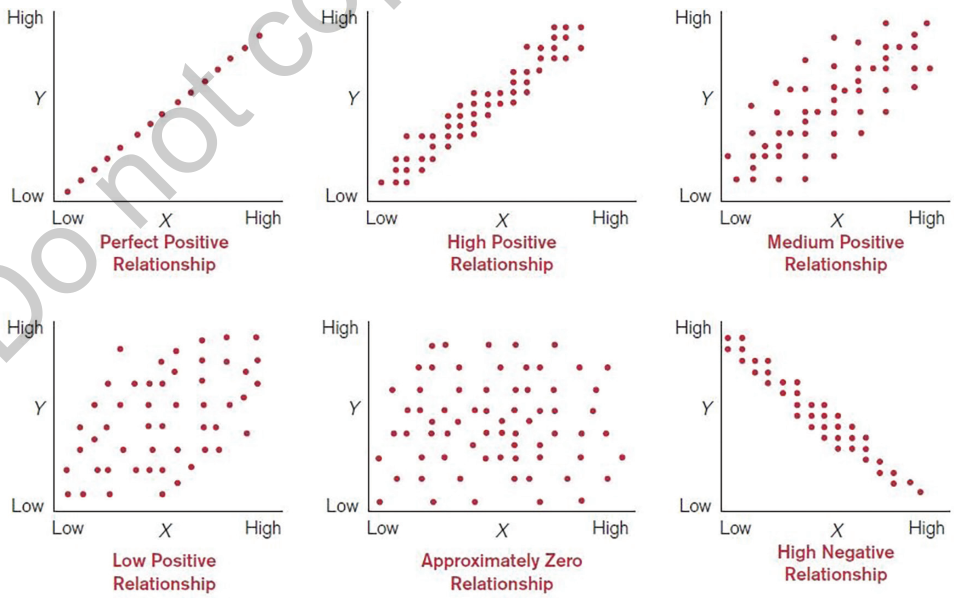
How to think about correlation numbers and graphs is summarized for you in Figure 2.3.

Caution! Correlation does not imply causation.

While it is *possible* that some correlations show causal relationships, be careful. The correlation could be caused by a third variable. In the case of a student who spends many hours studying and has a very good GPA, both of these outcomes might have been caused by the student's (1) motivation to do well, (2) level of pressure from parents, (3) amount of enjoyment of class subjects, or (dare we hope) (4) the skill and engagement of a fine professor.

FIGURE 2.3

Correlations always range from −1.00 to +1.00. The sign (positive or negative) indicates whether the two variables move in the same direction or in opposite directions. The number (from 0.0 to 1.0) tells you how well each data point fits onto a general pattern. If a correlation is zero, it means there is no pattern or association between the two variables.



SOCIAL PSYCHOLOGY IN POPULAR CULTURE

Correlation and Causation in *Harry Potter*


Pictorial Press Ltd / Alamy Stock Photo

Most people are familiar with the Harry Potter series of books and movies, in which a young boy discovers he's a wizard and attends a boarding school to learn spells. In the sixth book, *Harry Potter and the Half-Blood Prince* (Rowling, 2005), Harry first gets the chance to learn the magical power of "apparition," or the ability to disappear and reappear in a different location. Harry's first apparition lesson is an example of his awareness that correlation does not imply causation.

Harry describes his apparition teacher's appearance and how that appearance might be linked to the ability to disappear:

> He was oddly colorless, with transparent eyelashes, wispy hair, and an insubstantial air, as though a single gust of wind might blow him away. Harry wondered whether constant disappearances and reappearances had somehow diminished his substance, or whether his frail build was ideal for anyone wishing to vanish. (p. 382)

Harry doesn't know whether having a wispy appearance caused his teacher to have a greater apparition ability—or whether his talent at apparition has caused him to appear wispy. While Harry suspects there's a correlation or association between the two things, he knows that without more information, he can't know which is the cause and which is the effect. It's also possible that the association is merely a coincidence or that both are caused by something else. Without logical research designs, the secrets of apparition may remain unknown to Muggles (nonmagical folks) and wizards alike.

A weird example about ice cream will help you understand the "third variable" problem. There is usually a correlation between amount of ice cream sold per year in any given city and the number of people in that city who drown that year. Is ice cream consumption causing drowning or vice versa? Probably not. It's more likely that a third variable explains the correlation: heat. Towns that have hotter temperatures (such as Miami, Florida, and Austin, Texas) probably sell more ice cream. In addition, more people swim in these towns, also due to the heat (which unfortunately sometimes leads to more drownings). So, while it might look like the two variables of ice cream and drownings are related, both are actually caused by something else.

Even if there is a causal relationship, we might not know which is the cause and which is the effect. Think about the controversial idea that watching violence on TV causes children to act more violently in real life. The correlation seems to be real, but the causal relationship may go in the other direction. Maybe children inclined to be violent already are attracted to TV violence. Simple correlations are only clues. But they are a provocative way to begin engaging with research designed to solve social problems.

THE MAIN IDEAS

- The scientific method, which is used by social psychologists who conduct research, includes (1) observing a pattern, (2) generating a hypothesis, (3) scientifically testing the hypothesis, and (4) interpreting results so that the hypothesis can be refined and tested again.
- Three ways to gather descriptive data are (1) archival data, (2) naturalistic observation, and (3) surveys.
- Correlations test whether two continuous variables (or variables that have scores along a range) are associated with each other in a systematic way.

CRITICAL THINKING CHALLENGE

- Think about the classrooms you've been inside recently. Consider the physical aspects of the room (such as size, type of desks, color, art on the walls, and so on). Then consider how people choose to sit in the room during classes (such as whether they prefer the front or back row, how much they spread out, what kinds of people tend to sit together, and so on). Generate three hypotheses about how either the physical environment or the social environment shapes learning.
- Imagine that you want to do a study on how companies support leadership within their organizations. First, describe how you might conduct the study using archival data; then, how you'd do it with naturalistic observation. Finally, describe how you would conduct the study differently if you decided to give people who work there a survey. What kinds of questions would you ask? How would you get people to fill it out honestly?
- Describe two positive correlations you think are true in your own life, and identify two negative correlations you see in your own life. Then, choose one of those correlations and discuss whether you think there is a causal relationship between the two variables or whether a third variable drives what appears to be an association.

HOW DO EXPERIMENTS WORK IN SOCIAL PSYCHOLOGY?

>> LO 2.2: Understand the strengths and weaknesses of experiments and how their results are analyzed.

If the most famous book about experimental designs were going for a big audience in the self-help market, then it might be called *How to Think Clearly*. However, the helpful book written by Donald Campbell and Julian Stanley (1966) had a less dramatic title: *Experimental and Quasi-Experimental Designs for Research*.

The original target market for this groundbreaking book was the unruly world of education research. Campbell and Stanley (1966, p. 2) wanted to calm things down and remind researchers that experimentation takes time, replications, and multiple methods. They organized the world of experimental research design into three categories:

(1) preexperimental designs,
(2) quasi-experiments, and
(3) true experiments.

Preexperimental Designs

Main strength: everyone is treated equally. Main weakness: no comparison group.

The simplest experimental design is called a **preexperiment**. Here, a single group of people is tested to see if some kind of experience or treatment had an effect. Imagine a college is interested in making sure new students succeed, especially in their first year. The faculty design a class all incoming students have to take that's called something like "Freshman Seminar" or "University Success." The college requires *every* new student to take this class, then tracks that group's success through outcomes like their GPA at the end of the semester.

You can see that with preexperiments, because everyone is treated exactly the same, we avoid ethical concerns that might come up in research designs in which different groups are formed. For example, if the college required half of the incoming students to take the class and told the other half of students they weren't allowed to take the class, one of the groups might be at a disadvantage. This problem becomes even more clear when we think of research in areas like mental health interventions. If we design a new therapy to decrease anxiety, giving it to everyone who suffers from anxiety sounds the most fair.

But the problem with preexperiments also then becomes clear: Once we see results, can we really know any changes were because of the treatment? If grades in that year's incoming class are particularly good, how do we know they weren't just smarter than last year's class? And if anxiety decreases, was it the therapy—or something else that happened, like a change in the culture? We can be more confident that an experience or treatment has the effect we think it does when we use the next two research designs.

Quasi-Experiments

Main strength: enables comparison between groups. Main weakness: may not control alternative explanations.

Sometimes we want to compare groups that exist naturally. We might want to study people who have survived natural disasters (like tornadoes or hurricanes) compared to those who have never been in one. We might want to compare people who have served time in prison to those who haven't, people in the military versus civilians, athletes versus nonathletes, people who drink coffee versus those who don't . . . and so on.

Quasi-experiments compare outcomes between or among groups that are naturally occurring. For example, when comparing whether men or women are most likely to interrupt in a conversation, we use two groups (men and women) that existed regardless of our study. There are often extremely interesting questions that can only be asked through quasi-experiments, because people are already in their respective groups. We can now compare one group to another—which was impossible with preexperiments.

But here's the weakness: Even if we find a difference, we can't really be sure *why* that difference exists. In our interruption study, if one group interrupts more, is it because of genetics? Hormones? How boys versus girls are raised in our culture? Descriptive research (explained in the section above), preexperiments, and quasi-experiments all suffer from the same downfall: **confounding variables**.

Confounding variables are factors or issues that offer alternative explanations to why our results came out like they did, which limit our ability to ever say "Variable X caused changes in Variable Y." None of the methods of data collection described so far let us make claims about *causality,* because they all have confounding variables. The only way around confounding variables is a true experiment.

Preexperiment: A research design in which a single group of people is tested to see whether a treatment has an effect.

Quasi-experiment: A research design where outcomes are compared across different groups that occur naturally.

Confounding variables: Alternative explanations for why results came out as they did, which limit a researcher's ability to claim a causal relationship between variables in a study.

True Experiments

Main strength: controls (most) confounding variables. Main weakness: usually requires more time and effort.

The true experiment is the gold standard across all the sciences. That applies especially to social psychology because a **true experiment** compares two or more groups that are equal in every possible way except one (the variable we're testing in the study). The only way to make sure that the groups are equivalent to each other in every way *except* what we're studying is to use **random assignment**. Random assignment means that each person in the study has an equal chance of being put into any of the groups involved. We might flip a coin, draw names from a hat, or use a computer to assign people randomly to various experimental conditions.

Why does random assignment work—how does it eliminate confounding variables, or other explanations? With enough people in each group, random assignment creates a *statistical probability* that, on average, the people in each group are pretty similar to each other. If you have 100 people in your study and you know that half of them are men and half are women, then randomly assigning each person to be in either Group A or Group B of your study means there's a pretty good chance that Group A will consist of about 25 men and 25 women—and the same goes for Group B. Thus, random assignment creates roughly equal groups.

The idea is that in a true experiment with random assignment, the groups you're studying should be as equivalent as possible in every way *except* which group they're in. That way, if the two groups have different outcomes at the end of the study, the only possible reason must be the experimental manipulation—there's just no other explanation. Honestly, it is probably impossible to control every confounding variable. And random means just that—it's possible that random chance will lead the groups to be slightly different in other ways. But random assignment is the best technique we have, and it's why good *replications* are so important to any science.

True experiment: A research design comparing two or more groups that have been created with random assignment.

Random assignment: Placing participants into various conditions of a study using a chance method, to eliminate confounding variables by making the groups as equal to each other as possible.

Independent variable: A variable that is manipulated at the beginning of an experiment to determine its effect; it's how the groups are different from each other at the start of the study.

Independent and Dependent Variables

True experiments have two types of variables (see Table 2.1).

The **independent variable** creates comparison groups, based on random assignment, that will have different experiences in the study. It's the variable that's being manipulated by the experimenters. For example, researchers might have one group of

TABLE 2.1

Independent and Dependent Variables in Experiments

STUDY BASICS	INDEPENDENT VARIABLE	DEPENDENT VARIABLE
Students listen to either classical or rock music while they study, to see if music affects their memory on a test later.	Type of music (classical or rock)	Performance on the memory test
People write an essay about either death or puppies, then rate how much anger they feel.	Essay topic (death or puppies)	Level of anger
Children watch a commercial with dolls or with trucks, then are rated on how aggressively they play with clay and crayons.	Commercial topic (dolls or trucks)	Level of aggression
Sports fans see images of athletes wearing black jerseys or green jerseys and are asked to rate how well they expect each player to do that year.	Jersey color (black or green)	Expectations of players' performance

students listen to classical music and have the other group listen to rock music. So, the independent variable is *type of music*—it's the experimental manipulation that makes the groups different from each other at the very beginning of the study.

Of course, we also have to measure some kind of outcome, which is the **dependent variable**. Maybe while our two groups listen to either classical or rock music, everyone reads the same passage from a textbook. Then, we measure their memory on a test of the material: the dependent variable. In this case, we've hypothesized that *memory scores* are "dependent" on whether participants heard classical or rock music. In experiments, the independent variable is the *cause* being tested (here, type of music) and the dependent variable is the *effect* or outcome (here, memory test scores).

If the groups really were equivalent to each other in every way except the independent variable, any differences in the dependent variable must have been caused by which group they were in. This is how true experiments allow us to make causal inferences. Sometimes, studies include a neutral or baseline group as well that receives no treatment. For example, a study might include (a) a group of students who listen to rock music, (b) a group of students who listen to classical music, and (c) a group of students who listen to no music at all. In this case, that last group—the neutral group—is called a **control group**.

Dependent variable: The measured outcome at the end of an experiment that is affected by the independent variable.

Control group: A group of participants in a true experiment that serves as a neutral or baseline group that receives no treatment.

***t*-test:** A statistical test that compared the mean and standard deviations of two groups, to see if they are different from each other.

Analyzing Results in Experiments

Once you're done with the study and have your results, how do you know if the hypothesis was supported?

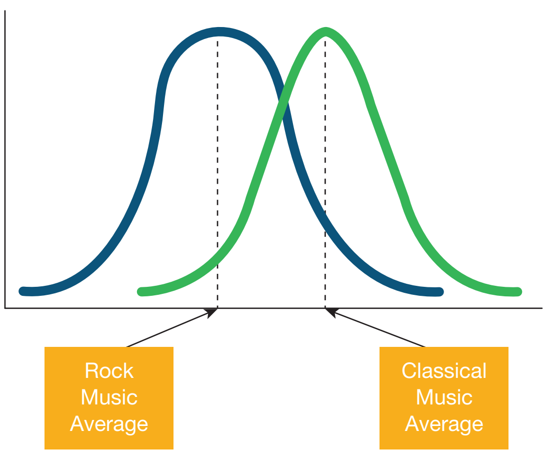
Comparing Two Groups: The *t*-Test Statistic

If your study only had two groups in it, you'll use a test called a *t*-test (see Figure 2.4).

The ***t*-test** statistic compares the mean and standard deviations of both groups to see if they're different from each other. You might find it interesting that this test was first created by a man named William Sealy Gossett, who used it to test whether different batches of Guinness beer tasted the same (Mankiewicz, 2000).

FIGURE 2.4

One way social scientists look for patterns is by comparing average scores within different groups of participants. When we compare two groups, as here, we use a *t*-test. When we compare three or more groups, we do an analysis of variance, or "ANOVA."



There were two critical components to Gossett's statistical invention: (1) the sampling had to be random, and (2) the sample had to be big enough to be representative of all casks of beer Guinness made in a given day. It would not be good for Guinness's business if the morning batch tasted different from the afternoon batch. Gossett's invention of the *t*-test statistic for comparing two groups (the morning vs. afternoon batches) ensured consistent quality in Guinness. When we use it for social psychology research, we usually want the groups to be *different* from each other, showing that our independent variable caused some change in the dependent variable. But we can still be grateful for how beer helped us reach our conclusions (thanks, beer!).

iStockPhoto/WaraJenny

We have Guinness to thank for the statistic known as the *t*-test.

Comparing Three or More Groups: Analysis of Variance

What if you have three or more groups in your study?

Think back to one of our earlier examples: Comparing memory test scores for students who studied while listening to rock music, classical music, or no music at all (the control group). Or, what if Gossett wanted to compare Guinness taste samples from more than two groups of beer—say, one from each day of the week?

The principle for comparing multiple groups is the same as comparing two groups. For each group, we calculate the mean and the standard deviation, just like before. However, when we're comparing three or more groups, the test is called an **analysis of variance**, or "ANOVA" for short. For example, perhaps the classical music group in a three-group experiment does better than the rock group—and maybe they are both better than the control group with no music at all. ANOVA tests will tell you whether at least one of the groups is different from the others. Different types of comparison groups require different statistical tests, and once you know the theory behind them, they can be fun to learn.

THE MAIN IDEAS

- A preexperiment is a method in which a single group of people is tested to see whether some kind of treatment has an effect.
- A quasi-experiment tests for the effect of a treatment in groups that have formed naturally, such as men versus women, or people who do or don't like to drink coffee.
- A true experiment compares outcomes on two or more groups that have been created by the experimenter through random assignment to condition. True experiments are the only methodology that can lead researchers to make claims about cause-effect relationships between variables. The variable that's manipulated and makes groups different at the start of the study is the independent variable, and the measured outcome at the end is the dependent variable.
- Researchers can compare the means and standard deviations of their various experimental groups to see if they differ from each other at the end of a study. When two groups are compared, a *t*-test statistic is used; when there are three or more groups, an analysis of variance test is used.

CRITICAL THINKING CHALLENGE

- Earlier, we discussed challenges to naturalistic observation (reactivity) and to surveys (social desirability). What are some challenges to experimental research designs? How could scientists overcome these challenges?

Analysis of variance (ANOVA): A statistical test that compares the means and standard deviations of three or more groups, to see if they are different from each other.

- Imagine that you have a hypothesis that people who drink a lot of caffeine will experience heightened emotions over the course of a day. Explain how you might test your hypothesis using a preexperiment, a quasi-experiment, and a true experiment.
- True experiments are needed to test things like advances in medicine. If someone found a pill they think could cure cancer, a true experiment would require a study in which participants with cancer are randomly assigned to either a group that gets the pill or a control group of people who get a placebo. Neither can know whether they're getting the "real" pill or the placebo. What are the ethics of this kind of study, where people's lives might actually be in the balance?

HOW CAN I RECOGNIZE TRUSTWORTHY RESEARCH?

>> LO 2.3: Explain why reliability, validity, random sampling, ethics, and open science signal good research.

You are asked to consume research many times each day.

Some of it is legitimate, but much of it is nothing more than "I read somewhere . . ." Research claims on TV or online about foods, drugs, and toothpaste are uncertain. Advertisements clearly have a conflict of interest in whether they're honest about the benefits of their products. Recognizing trustworthy friends, products, and research requires similar skills built from painful life experiences. You may not develop those skills until you've been burned a few times.

This section answers the core question "How can I recognize trustworthy research?" by describing

(1) the meanings of reliability, validity, and random sampling;
(2) the ethical guidelines that support social psychological research; and
(3) the "open science" movement.

Shutterstock/Andrey_Popov

If you use a scale like this one to weigh yourself, how confident are you that the result is correct or that it wouldn't change if you stepped off and back on?

Reliability, Validity, and Random Sampling

Part of the appeal of digital textbooks is practical: Printed textbooks are heavy!

Sometimes, your arms and back hurt from lugging them around all day. So, you become curious about just how much those suckers weigh. You pile them on top of the scale at your local gym and it gives you a result: 36.8 pounds.

No wonder your back hurts! But then your friend tells you that the scale is often off by several pounds, which she knows because she sometimes compares what that scale tells her to the scale at her doctor's office on the same day. So you try again. This time the same scale tells you that the exact same pile of books weighs 33.2 pounds. Surprised, you take the books off and put them back on, and now it says 40.6 pounds! Clearly, you can't trust this scale because it is unreliable.

It doesn't matter if we are research psychologists, clinical psychologists, or just trying to have an intelligent conversation. If we run into the same problem weighing self-esteem instead of textbooks, then we don't really know what to believe about self-esteem.

Reliability is consistency of measurement, over time or over multiple testing occasions. It doesn't matter if it is a weight scale, an intelligence test, a train that does not keep to schedule, or a friend who may or may not show up on time: To be trustworthy, consistency is key.

A study must also be high in internal, construct, and external validity. High **internal validity** signals confidence that your results mean what you think they mean. For example, if you forgot to use random assignment in the music and memory study, then you have opened the door to self-selection and alternate explanations. Randomly assigning participants to different conditions reduces that threat to internal validity because it helps eliminate confounding variables. Another way to improve internal validity is to ensure **construct validity**, which is whether the tests, surveys, and so on chosen for the study really measure what we think they're measuring.

High **external validity** builds trust in your results if they apply to other people or settings. For example, many social psychology studies have used college students as participants. That's great—if you are a college student and the results apply to you. But those results might be limited in their **generalizability**, or how much they apply to people "in general." This problem is sometimes known by the acronym WEIRD because so many participants are Western, Educated, and from Industrialized, Rich, Democratic cultures. We know less about other kinds of people—the diversity in participants is sometimes lacking.

One of the best ways to increase generalizability—and thus external validity—is to use **random sampling** (not the same thing as random assignment). Random sampling means that the people in your study were randomly selected from the larger population of interest. If you want to know the opinions of students at your college or university, the easy approach would be asking for volunteers or surveying people who live in a certain dorm. But if you want data you can trust, random sampling would mean getting a list of every student and choosing the people in your study in some truly random way. Hopefully, that technique results in a sample of participants who represent the diversity of people across your entire population.

Ethics and Institutional Review Boards

"The therapists . . . they should be in prison right now, today."

It is distressingly easy to lose our moral compass. Bob Kelly had good reason to be critical of therapists during a presentation at the Duke University Law School. As owner of the Little Rascals daycare in Edenton, North Carolina, Bob (and six others) had been falsely accused of sexually abusing some 90 children. There was no physical evidence—none!

So, some of the most important "evidence" included reports from psychological therapists. Sentenced to 12 consecutive life sentences, Bob Kelly's conviction was overturned—but only after 6 years in prison. It was also unlikely that the therapists had lied. They had somehow lost sight of the ethical guidelines that guarded them and their profession. It seems they genuinely thought he was guilty, even though there was no evidence of that conclusion. We have a solemn responsibility to treat people with respect. That includes how we use unobtrusive methodologies like naturalistic observation and even archival studies when we examine people's life information.

Researchers (across the sciences) rely on **institutional review boards (IRBs)**, committees that consider the ethical implications of any study. *Before* any of us begin formal research that might affect participants and be published, we are obliged to submit our methods for review. Your local IRB committee is typically composed of representatives from different departments in a college, university, research institute, or corporation—and sometimes from a combination of such organizations.

Many committees include a lawyer and sometimes a member with no background in research at all, to represent the "average person's" perspective. The IRB committee will consider applications for every study your university might run, then decide in

Reliability: Consistency of measurement, over time or multiple testing occasions. A study is said to be reliable if similar results are found when the study is repeated.

Internal validity: The level of confidence researchers have that patterns of data are due to what is being tested, as opposed to flaws in how the experiment was designed.

Construct validity: The degree to which tests, surveys, and so on chosen for a study really measure what we think they're measuring.

External validity: The extent to which results of any single study could apply to other people or settings (see *generalizability*).

Generalizability: How much the results of a single study can apply to the general population (see *external validity*).

Random sampling: A sampling technique where a researcher randomly chooses people to participate from a larger population of interest.

Institutional review boards (IRBs): Committees of people who consider the ethical implications of any study before giving the researcher approval to begin formal research.

The National Archives at Atlanta

Ethical violations in the Tuskegee Syphilis Study helped establish the need for IRBs to prevent future violations of trust. Participants in this study of syphilis were not (a) told the purpose of the study, (b) given a chance to leave the study, or (c) treated with penicillin that might have cured them.

advance if there is any potential serious harm involved. If there is possible danger, the study won't be approved.

The specific guidelines for psychology come from both our local IRB committees and the **American Psychological Association (APA)**, an organization of professionals in the field who have determined what the ethical standards of research should be. Note that the standards currently in effect from the APA were not always in place. Some of the studies you'll read about in this book were done before APA standards and IRB committees were standard—and thus, some of those studies have been harshly criticized for being harmful to the people who were in them. Protecting and respecting participants should be the highest priority for anyone conducting research. What are the guidelines?

American Psychological Association (APA): A large organization of professional psychologists who provide scholarly publications, writing guidelines, and ethical standards for research.

The American Psychological Association lists several "rights" that they say all participants should have. Again, consider each of these as you read about studies in this book and consider whether you think the study was ethical. In addition, if you have the chance to participate in any studies yourself, it's important for you to know what your rights are—so that you can stop participating at any point if you are uncomfortable or

SPOTLIGHT ON RESEARCH METHODS

Questions Your Institutional Review Board Might Ask

Institutional review boards (IRBs) are committees that review ethical considerations for any new study; you need to get IRB approval before you can move forward. To know more about this process, consider the questions below; they are the typical types of questions any researcher will need to answer before starting the research process. Researchers are required to type their answers and submit them to the IRB committee for review, along with copies of any materials such as surveys, videos, and so on.

- What is the purpose of your study, and what are the hypotheses?
- What kinds of participants do you plan to use? How many? How will you recruit them? Will they be compensated for their participation through something like money or extra credit?
- Are the participants from any kind of legally protected group, such as children, prisoners, people with disabilities, and so on?
- How will you get informed consent?
- What are the specific questions you will ask or experimental procedures you will use?
- Will there be any deception? If so, what is the justification for that deception?
- How will debriefing occur?
- What are the potential harms (physically, emotionally, or mentally) that participants might experience, short term or long term?
- Will you provide any resources if participants have questions or concerns?
- Are there any potential benefits participants might experience as a result of being in your study?
- How long will your study take? How will you ensure that the data collected will remain confidential and/or anonymous?
- Do you plan to present your results to the public, such as through a conference presentation or publication? If so, will the participants have access to these results themselves if they are interested?

Hopefully, you can see that IRB committees take their jobs very seriously—as they should. There also are career opportunities waiting for you if you are interested in protecting the rights of both human and nonhuman animals. These careers combine elements of law, philosophy, science, and psychology.

feel like you don't want to continue for any reason. Some of the participant rights identified by the APA are as follows:

- **Informed consent**: Participants should be told what they will be asked to do and whether there are any potential dangers or risks involved in the study before it begins.
- **Deception**: Participants should be told the truth about the purpose and nature of the study as much as possible. *Deception,* or hiding the true nature of the study, is only allowed when it is necessary because knowing the truth would change how the participants responded.
- **Right to withdraw**: Participants have the right to stop being in the study at any time, for any reason, or to skip questions on a survey if they are not comfortable answering them.
- **Debriefing**: After completing the study, all participants should be given additional details about the hypotheses of the study, allowed the opportunity to ask questions, and even see the results if they wish. This *debriefing* after the study is complete should definitely include an explanation of any deception that was involved (if deception occurred) so that participants have the right to withdraw their data if they are upset about the deception.

If you want to design your own study, you should consider all these criteria for the quality of good research. In addition, you'll have to get approval from your school's IRB committee as well before you begin. To learn more about the IRB approval process, see the Spotlight on Research Methods feature.

The "Open Science" Movement

Ethics are always important.

We've already discussed some ethical considerations, such as avoiding deception in studies whenever possible, making sure we get informed consent before participants start in a research project, and so on. The ethics of science are even broader, though, when we start to think about how studies happen from start to finish. What if a researcher misrepresented their results or decided to form hypotheses only after they knew how the study's results turned out? What if they refused to share their data with other people, who could confirm the findings?

Open science is a movement to make scientific research transparent, accessible, cooperative, reproducible, and honest. The aim of open science is to remove barriers for the creation of studies, sharing of data and results, and analysis of implications or conclusions. It's a way of saying to others, let's all do this together in an open, honest environment. Open science is also a reaction to the "replication crisis" discussed in Chapter 1. As scientists, social psychologists want to be honest and confident about our research conclusions.

There are several ways that open science encourages this kind of communication and exchange; a few are preregistration, results-blind peer review, and publication badges.

Preregistration

Imagine that a scientist does a study in which they're not really sure what they're looking for.

This is called exploratory research, and there's nothing wrong with it. But now imagine that after the results are analyzed, the scientist publishes the study and more or less pretends that they predicted the outcomes from the beginning. They look super smart! But it's not an honest approach.

Open science's solution is **preregistration**, a practice of specifying—in advance—your hypotheses, procedure, and statistical plan for analyses (see Nosek et al., 2017).

Informed consent: Participants' right to be told what a study will involve, including potential dangers, before the study starts.

Deception: Hiding the true nature of an experiment from a participant so they act more naturally.

Right to withdraw: The right participants have to stop being in a study at any time or to skip questions on a survey.

Debriefing: Additional details given to participants after participation in an experiment.

Open science: A movement to make science more transparent, cooperative, reproducible, and honest.

Preregistration: Specifying your hypothesis, procedure, and statistical plan for a study before collecting data.

This plan is made publicly available to anyone, so you are committing to everything in an open, transparent way before you begin your study. Several preregistration templates have been created to help people through this process, where researchers can post their plans on independent websites.

Preregistration is not without problems. For example, you might say that you're going to get 100 people for your study, but you can only get 75. Or you might assume that people will pay attention to instructions during your procedure, but some of them don't and they mess up what they are supposed to be doing. Or you might realize after you've collected data that you had a typo on your scale that changed what the question was asking. Scientists are certainly not perfect, and mistakes can be made. But all of these changes can simply be documented and explained. That way, readers of the research can understand exactly the process that occurred and why changes had to be made.

Typical questions you'll answer on a preregistration form are things like this:

- What are your hypotheses? If you're doing a quasi-experiment or experiment, what are the independent and dependent variables? If you're doing a correlation study, do you expect a positive or negative correlation?
- What exactly will the procedure be—what will participants do? What will be the order of procedural steps? How long with it take each person to do the study? How will you do random assignment (if relevant)?
- How will you recruit participants, and how many do you expect to find? Will anyone be excluded from data analysis—and if so, why?
- How will each of your variables be measured and calculated? What statistical tests will you use to analyze the results?

Results-Blind Peer Review

Every academic field has professional journals, where researchers publish their results.

Most of these journals are what we call "peer-review" journals. That means that before any article is accepted for publication, it's sent out to other experts on that topic to see what they think. Those people, called reviewers, give the author(s) anonymous feedback about whether they think the article is worthy of being published. Sometimes the reviewers will make suggested changes that they want to see, and if those changes are made, the article will be published. Sometimes, however, the reviewers can simply say that they don't like the study and stop it from being published.

Until the open science movement, all of this reviewing happened after a study was completed and written up. That meant that the peer reviewers knew how the study turned out. The problem with this is that it can lead to biases in what is and isn't accepted for publication. Maybe the reviewers wouldn't like the results because they go against a theory they favor. A more common problem is that studies usually weren't published if their results didn't show statistically significant findings or results that matched their hypotheses.

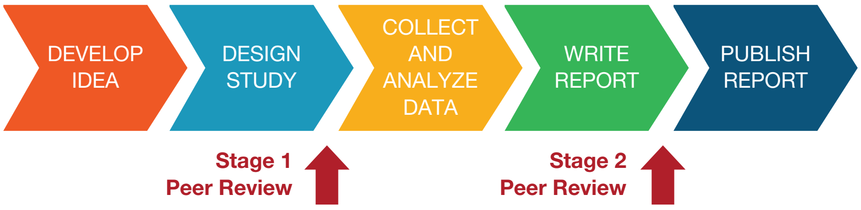
These problems can largely be eliminated with a practice called **results-blind peer review**, which means that reviewers are asked about the importance of the study *before* they see the statistical outcomes, as shown in Figure 2.5. If they agree that the study has merit, they accept it for publication at this stage. Reviewers will also be asked for their feedback after the results are calculated—but now they comment on whether the study followed the preregistration plan and interpreted everything correctly.

That way, even if the results surprise everyone, the study still gets published. Chris Chambers, the chair of a committee at the Center for Open Science, stated the benefits of this process like this: "The incentives for authors change from producing the most beautiful story to producing the most accurate one" (Center for Open Science at https://cos.io/rr/). Just like a relationship partner, science is even more beautiful when it's accurate and honest.

Results-blind peer review: Asking experts to judge a potential study's value and quality before the data have been collected and analyzed.

FIGURE 2.5

When an article goes through the results-blind peer review process, outside experts give feedback about the quality and importance of an article before the data are actually collected. Then, they review a second time, focusing on whether the study followed the original design plan.



Publication Badges

Beyond the rewards of knowing you're doing good science, what incentives are there for people to engage in open science practices?

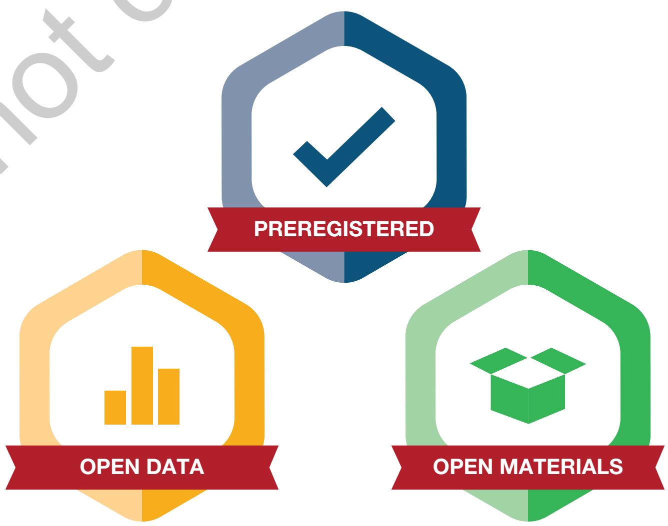
One reward is the use of **badges**, or visual icons that mark a study with signals that it has followed these procedures. You can see what the badges look like, at least for some journals, in Figure 2.6. If a study was preregistered, once it's published, the red badge will appear on the first page. If the authors have made their procedural materials available to everyone, they get the orange badge, and if they have posted their original, raw data spreadsheets online, they get the blue badge.

Over 50 journals now use the badge system, and early trends show that they really do increase the number of scientists who share their data publicly (Kidwell et al., 2016; Rowhani-Farid et al., 2017). The open science movement is likely going to increase in

Badges: Visual icons that can mark if a study used open science practices.

FIGURE 2.6

Professional journals are increasingly marking studies with these images, called badges, when they follow open science guidelines. These examples are from the Center for Open Science.



usage and popularity over the next several years, as many people see it as the only way to make the scientific process truly objective and transparent.

THE MAIN IDEAS

- Reliability, validity, and random sampling are all criteria regarding the quality of any given research study.
- Ethical considerations are also very important when evaluating research studies. The American Psychological Association lists several participant rights, such as informed consent and debriefing. In addition, any study must be approved by an IRB before it can be conducted.
- "Open science" is a movement to make research more transparent, cooperative, and honest. It involves preregistration, results-blind peer review, and badges indicating whether studies have used the process before publication.

CRITICAL THINKING CHALLENGE

- Imagine a study conducted in 1930, before the APA enacted ethical guidelines and before IRB committees were common. If that study were unethical but highly interesting in terms of the results, should textbooks still talk about the study and what we learned from it? Does continuing to talk about the study disrespect the participants, or does learning from it mean that at least we are attempting to get some good from the bad that already occurred?
- Different IRB committees have different levels of strictness regarding ethical thresholds. For example, one committee might be fine with a study that causes participants to temporarily be angry, sad, or aggressive—while another committee might consider the same study unethical. If you were to serve on an IRB committee, how would you decide what the threshold of danger or harm should be? What's the balance between possible risk of harm versus what could be learned from the study?
- Some professional journals charge for copies of their articles or require people to pay for subscriptions. Others offer their articles to readers for free, but they require that the scientists themselves pay to publish their work in the journal. What do you think is the best system for research to be available to other scientists or the general public, in terms of how it is funded? Should there be a new system, like a "science tax" that everyone pays but is used to make scientific progress available to everyone? Discuss how you think science should be funded and why.

CHAPTER SUMMARY

How Do Social Psychologists Design Studies?

Basic researchers advance theories, while applied researchers translate those theories into real-world settings or people. Both types of researchers use the scientific method, a systematic way of creating knowledge. Descriptive designs clarify patterns of data without experimenter intervention. Examples of descriptive designs are use of archival data, naturalistic observation, and self-report surveys. Each method has strengths and weaknesses. Descriptive data are often analyzed using correlation tests. Correlations tell whether two continuous variables have a system association with each other but cannot tell whether there is a causal relationship between those variables.

How Do Experiments Work in Social Psychology?

Preexperiments are studies in which a single group of people is tested to see if a treatment or experience has an effect. When scientists want to compare two or more groups to each other, they can use quasi-experiments or true experiments. Quasi-experiments compare naturally occurring groups to each other, while true experiments create groups using random assignment (participants are randomly put into one of the study's groups). The variable that makes groups different from each other at the start of a study is the independent variable, and scientists are testing whether that has an effect on the outcome of the study, which is called the dependent variable. To analyze results, a *t*-test is used when there are only two groups in a study, while an analysis of variance (ANOVA) tests for differences in three or more groups.

How Can I Recognize Trustworthy Research?

Reliability, validity, and random sampling are all important things to consider when trying to evaluate the quality of a research study. Reliability is whether the measures used are consistent. Internal validity is the extent to which results are interpreted in an accurate way, based on the setup of the study. External validity is the extent to which results could apply to other people or settings. Random sampling will hopefully result in a diverse variety of different kinds of people in the study. In addition, the ethics of any proposed study will be reviewed by a committee called an institutional review board (IRB). Examples of ethical considerations are whether a study includes informed consent, deception, and debriefing. Finally, the "open science" movement is a trend in psychology that encourages researchers to be honest, open, and transparent with their research from start to finish—for example, they might post spreadsheets of their data online so that other scientists can check the accuracy of their statistical analyses.

CRITICAL THINKING, ANALYSIS, AND APPLICATION

- Which of the research methods described in this chapter seem the most appealing to you? Why is that method appealing, and what issues or concerns do you have with the other methods?
- Find a news report that makes a claim that one variable causes another (for example, "Drug X leads to bad behavior," or "Access to birth control leads to risky sex," etc.). Is the causal relationship being suggested one that seems valid? Why or why not? How could this causal claim be scientifically tested?
- How many times does a study have to be replicated for researchers—or the general public—to be confident in the results? Even if a study is never replicated, does that mean that the data are useless? What other explanations could there be for results that seem to change (in other words, aren't replicated)? Is it possible that researchers simply haven't identified the exact reason why the pattern happened the first time, for example?
- Ideally, studies have generalizable samples of people who participate. But what if you want to generalize your findings to all of humanity? It's clearly impossible to use random sampling across everyone in the world, so most people just use participants who are nearby volunteers. What's the balance between convenience for the researcher and the need for a diverse, generalizable sample of participants? Are any studies truly high in external validity if they don't have true random sampling?